

From paper-and-pencil to screen-and-keyboard : studies on the effectiveness of internet-based marketing research

Citation for published version (APA):

Deutskens, E. C. (2006). *From paper-and-pencil to screen-and-keyboard : studies on the effectiveness of internet-based marketing research*. [Doctoral Thesis, Maastricht University]. Universiteit Maastricht. <https://doi.org/10.26481/dis.20060120ed>

Document status and date:

Published: 01/01/2006

DOI:

[10.26481/dis.20060120ed](https://doi.org/10.26481/dis.20060120ed)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

FROM PAPER-AND-PENCIL TO SCREEN-AND-KEYBOARD:

STUDIES ON THE EFFECTIVENESS OF
INTERNET-BASED MARKETING RESEARCH

© 2006, Elisabeth C. Deutskens

All rights reserved. No part of this publication may be reprinted or utilized in any form or by any electronic, mechanical or other means, no known, or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without written permission from the copyright owner

ISBN: 90-9020091-6

Printed by: B+W Computerformulare

FROM PAPER-AND-PENCIL TO SCREEN-AND-KEYBOARD:

STUDIES ON THE EFFECTIVENESS OF
INTERNET-BASED MARKETING RESEARCH

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van de Rector Magnificus, Prof. Mr. G.P.M.F. Mols
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op vrijdag 20 januari 2006 om 12.00 uur

door

Elisabeth Christine Duitskens

Promotoren

Prof. Dr. J.C. de Ruyter

Prof. Dr. M.G.M. Wetzels

Beoordelingscommissie

Prof. Dr. J.G.A.M. Lemmink (voorzitter)

Prof. G. Albaum, Ph.D. (University of New Mexico, USA)

Prof. Dr. R.A. Roe

Für meine Eltern

Acknowledgements

After finishing the four-year PhD program, I came to the conclusion that the challenge in writing a doctoral dissertation does not lie in the content but in the process. On my journey I was accompanied by several people who greatly supported me along the way.

First of all, I would like to thank my supervisors Ko de Ruyter and Martin Wetzels. The combination of Ko's conceptual strengths and Martin's methodological skills make them an unbeatable team and proved to be a huge and invaluable knowledge base for me. In addition to content related support, they significantly charted my route by giving me the great opportunity to put my ideas into effect. They allowed me to take responsibility and supported my interest in venturing outside the program by encouraging me to spend one semester abroad and to attend several conferences and colloquia. Thinking back, I also recall numerous social occasions that were just fun. Thanks for your help and support along the way.

In addition, I would like to thank Jos Lemmink, not only in his function as the head of my dissertation committee, but especially because of his time as head of our department where he created a productive and collegial atmosphere. I also thank Jerry Albaum and Rob Roe for valuable feedback on my dissertation.

Furthermore, I am indebted to Niels Schillewaert from InSites for giving me the opportunity to collect data by utilizing their online panel. Thanks also to Paul Oosterveld and Pieter Willems from Millward Brown for their financial and collegial assistance and Sonja van den Berg for her help with the data collection for the first study.

I would like to thank everyone in the marketing department for the great work environment. Special thanks go to Mark for his e-support, Nicole for her editorial help, and Karolina, Jacqueline, Jeroen, and Simon for proof-reading parts of my dissertation. I am indebted to Ad de Jong for his invaluable help with the generalizability analysis. I am especially grateful to Jill for extremely thoughtful and constructive feedback and for being a friend. Most of all, I would

like to thank Caroline and Vera. From day one, they took me under their wing and helped and supported me wherever possible. But above all, I would like to thank them for their friendship.

I would also like to thank the 2003 class of the Oxford Internet Institute Summer Doctoral Programme and especially Zoe Dowling for editing my first study. I spent a great semester at ASU thanks to the friendliness and hospitality of all faculty, staff, and PhD students. Special words of thanks go to Carolyn, whom I hold in high regard for her generosity and strengths.

I would also like to thank my friends for providing welcome distractions from work. I am also extremely grateful to my family, in particular my grandmother, my godchild Lydia, my godmother Brigitte, and my godfather Auwi. Special thanks go to my siblings, Annette and Christoph. Thank you for being the special sister and brother you are and for standing by my side through the years. I am most grateful to my parents. Their love and support are the basis of everything and I want to thank them by dedicating this dissertation to them.

Finally, I want to thank Alexander, my partner in life. His strength, humor, encouragement, help and support, and above all his love, mean everything to me and were the greatest assistance throughout the past four years. Thank you so much!

Table of Content

Chapter 1: Introduction.....	1
1.1 General Objective	2
1.2 Contribution	3
1.3 Development and Importance of Marketing Research	4
1.4 From Paper-and-Pencil to Screen-and-Key-board	5
1.5 Effectiveness of Online Marketing Research	10
1.6 Objectives of the Different Studies	11
1.7 Outline of this Dissertation	14
 Chapter 2: Response Rate and Response Quality of Internet-based Surveys: An Experimental Study	 17
2.1 Introduction	18
2.2 Literature Review and Hypotheses Development.....	19
2.3 An Empirical Study.....	23
2.4 Results	25
2.5 Discussion and Implications.....	31
2.6 Limitations and Future Research Guidelines	33
 Chapter 3: An Assessment of Equivalence between Online and Mail Surveys in Service Research.....	 35
3.1 Introduction	36
3.2 Literature Review.....	37
3.3 An Empirical Study.....	43
3.4 Analyses and Results.....	44
3.5 Discussion	48
3.6 Limitations and Future Research Guidelines	50

Chapter 4: Comparing the Generalizability of Online and Mail Surveys in Cross-National Service Quality Research.....	53
4.1 Introduction	54
4.2 Literature Review	55
4.3 Research Setting.....	58
4.4 An Empirical Demonstration	62
4.5 Discussion	74
4.6 Limitations and Future Research Guidelines	75
 Chapter 5: Individual Differences in motivation to participate in online surveys and the effect on response rate and response quality	77
5.1 Introduction	78
5.2 Conceptual Development of the Survey Participation Inventory (SPI)	80
5.3 Research Design and Methodology.....	89
5.4 Discussion	103
5.5 Theoretical and Managerial Implications	106
5.6 Limitations and Future Research Suggestions.....	107
 Chapter 6: Conclusion	109
6.1 Synopsis.....	110
6.2 Discussion	111
6.3 Perspectives on Future Research	116
 Appendix	119
 References	123

CHAPTER 1: INTRODUCTION

This chapter starts with a brief introduction to the topic and the main contribution of this dissertation. Next, it continues with a description of how marketing research developed and its relevance for today's business environment. Then, the advantages as well as disadvantages of online surveys and the different forms of online research are discussed in detail. Afterwards, the purpose, research questions, and intended contributions of the separate chapters of the dissertation are presented. Finally, the outline of this dissertation is explained.

*"Online research is more than just a new way to gather information,
more than just a new modality of research.
It represents a technological and cultural change.
What we do as researchers and how we think about research
are changing because of what we can do online."
(Miller and Dickson 2001, p. 140)*

1.1 General Objective

Online¹ marketing research is booming and presents the largest growing segment of the research industry. ESOMAR, the World Association of Research Professionals, reports that online surveys account for 35% of the US survey research market by the end of 2004 (ESOMAR 2004a) and Metzke and Allan (2005) predict an increase to 39% in two years time. Similarly, Burke, a premier international research and consulting firm verifies an increase from \$3 million spent on commercial online research efforts in the US in 1996 to half a billion at the end of 2001 (Hogg 2002). Online surveys focus no longer on the Internet as a research subject but are reflective of the general market research mix (Metzke and Allan 2005). Also academics increasingly use the Internet to collect data (e.g., Iqbal, Verma, and Baran 2003; Lynch and Ariely 2000; Mandel and Johnson 2002; Meuter et al. 2000; Novak, Hoffman, and Yung 2000; Toubia, Hauser, and Simester 2004).

Despite the increasing popularity of online research, empirical studies on the design and quality of online research are scarce. Some authors have started to test the influence of certain format or design parameters (Couper, Traugott, and Lamias 2001; Dillman et al. 1998; Lozar Manfreda, Batagelj, and Vehovar 2002; Sheehan and McMillan 1999). Several other studies have attempted to examine the quality of online surveys in comparison to traditional mail surveys (e.g., Epstein et al. 2001; Klassen and Jacobs 2001; Knapp and Kirk 2003; Shermis and Lombard 1999; Stanton 1998). Yet, these studies provide inconclusive and contradicting results. Also online access panels, which are very popular and growing steadily, have received little attention.

¹ "Online" typically refers to being connected to the Internet. The "Internet" stands for the network while the "Web" is just one service on the Internet. Yet, these terms are often used interchangeably. Since all three terms are common in the survey literature, online/Internet/Web survey will be used synonymously throughout this dissertation. In online marketing research, respondents are typically invited by an e-mail that includes a hyperlink to a web-questionnaire.

These issues essentially relate to the effectiveness of online surveys. In contrast to efficiency where the focus is purely on the resources spent, effectiveness implies the capability of, or success in, achieving a given research objective. Thus, effectiveness requires, beyond cost efficiencies, high quality data that fulfill specified research objectives. However, evidence on the cost structure of online surveys is largely anecdotal and empirical studies on the quality of online surveys are scarce. Therefore, this dissertation aims to enhance the evaluation of the effectiveness of online surveys by conducting four consecutive studies that examine different aspects of effectiveness. The central problem statement of this dissertation is:

What are the critical factors that influence the effectiveness of online marketing research?

1.2 Contribution

This dissertation contributes to our knowledge of research methods in several important ways. Since the virtual environment of online surveys has changed the surveying process, results with regard to offline surveys are not per se generalizable to the online setting. For example, online surveys differ from traditional research methods with respect to the method of contacting, the medium, and the mode of responding (Tourangeau, Rips, and Rasinski 2000). Therefore, it is likely that the interactive and impersonal nature of the Internet causes differences in the quality of online surveys. Despite the increasing popularity of online surveys, little methodological research has thoroughly examined the influences of various factors on response rates as well as response quality in online surveys. Little research on survey methodology has examined both B-2-B and B-2-C populations. Also, most previous studies have methodological limitations, are often narrow in scope, and limit themselves to the assessment of response rates only. To fill the lacunae regarding the quality of online surveys, this doctoral dissertation aims at gaining deeper insights into which factors matter and how they influence the quality of Internet-based research. Market researchers can benefit from the findings of this dissertation since specific suggestions for the design of online surveys and the use of online panels are offered.

1.3 Development and Importance of Marketing Research

Before continuing with detailed information on the different studies in this dissertation, the development and importance of (online) marketing research is illustrated. In the beginning of the 20th century the demand for consumer goods largely exceeded supply. Manufacturers could easily sell their products and thus did not spend much time thinking about customer orientation or marketing strategies. Consumers neither had much choice nor power, as illustrated by Henry Ford's famous quote that Americans could have any color car they want as long it is black. Yet, major technological advances in the following decades improved production drastically. Due to increasingly saturated markets and efficiencies in production, competition increased significantly so that nowadays consumers have a wide variety of products to choose from. This increase in competition has led to increased customer power since they can select from a large variety of products and services whatever satisfies their needs best. To gain a solid marketing share, companies have to be market oriented and produce what people need and want. In the late 1980's, seminal articles by Kohli and Jaworski (1990), Narver and Slater (1990), and Day (1994) started to shed more light on the concept of market orientation, which is defined as "organization wide generation, dissemination, and responsiveness to market intelligence" (Kohli and Jaworski 1990, p.3). Thus, market orientation involves multiple departments sharing information about customers and engaging in activities designed to meet customer needs. Moorman and Rust (1999) take this idea further by suggesting that the chief responsibility of the marketing function is to provide information about the connection between the customer and the product, service delivery, and financial accountability. This view entails that companies develop products based on customer needs and measure their level of customer satisfaction. Knowledge about customers can be gained through effective, highly used formal and informal marketing information systems. One essential pillar of marketing information systems is marketing research where customer satisfaction, repurchase intentions, brand equity, and advertising and promotion exposure and effectiveness are measured. The American Marketing Association defines marketing research as *"the function that links the consumer, customer, and public to the marketer through information--information used to identify and define marketing opportunities and problems; generate, refine, and evaluate marketing actions; monitor marketing performance; and improve understanding*

of marketing as a process. Marketing research specifies the information required to address these issues, designs the method for collecting information, manages and implements the data collection process, analyzes the results, and communicates the findings and their implications" (American Marketing Association, <http://www.marketingpower.com/mg-dictionary-view1884.php?>). Key in this definition is information - information about what customers need, information about target markets and marketing strategies, and information about whether products or services satisfy customer needs. The role of marketing research is to create this information.

1.4 From Paper-and-Pencil to Screen-and-Keyboard

The growing need for information from consumers about what they want has led to the development of an 18,922 million dollar marketing research industry (in 2003) that is still booming. According to ESOMAR, the world association of research professions, global market research spending grew by 5.1% in 2003 (ESOMAR 2004b). Product or concept testing, customer satisfaction studies, attitude and usage studies, brand share, advertising effectiveness and many more areas are surveyed. Yet, the industry is threatened by rising costs as response rates remain at or near record lows (Morin 2004). Telephone surveys for example are afflicted with growing time pressure, rise of telemarketing, more mobile phones and less fixed lines, answering machines, caller ID, call blocking and no answers (Morin 2004). Mail surveys suffer from poor response rates, slow response and manual transcription of data (Illieva, Baron, and Healey 2002). Furthermore, Wind (1997) indicates seven additional areas of discontent: (1) perceived lack of appropriateness (2) lengthy research cycles, (3) questionable validity of findings (4) questionnaire complexity/oversimplification, (5) a domestic focus, (6) non-user friendly data collection and analysis, and (7) restrictions on research involvement. Also McDonald and Adams (2003) ascertain falling participation rates, rising costs, respondent fears concerning misuse of personal information, and managerial issues resulting from the time taken to conduct postal surveys.

Parallel to the dissatisfaction with existing survey methods, another trend emerged. Researchers could observe a growing number of Internet users, reaching 934 million by the end of 2004. The Computer Industry Almanac projects that the threshold of one billion is crossed in 2005, reaching 1.35 billion

users by 2007 (www.clickz.com/stats/sectors/geographics/article.php/151151). Triggered by the growing number of Internet and e-mail users, a new research mode emerged: Internet-based research. Online marketing research is a new avenue for businesses to explore the world of their customers.

1.4.1 Advantages and Disadvantages of Online Surveys

Before the different types of online surveys and the key differences to traditional modes are explained in more detail, this section focuses on the advantages and disadvantages of using online surveys.

Compared to other research modes, online surveys offer a couple of key advantages (e.g. Green, Johnson, and Neal 2003; Illieva, Baron, and Healey 2002). First of all, online surveys have a very appealing cost structure since printing, postage, and data entry expenditure are eliminated. Even though the fixed costs of online surveys can be slightly higher because of the programming efforts upfront, variable costs are virtually zero and thus online surveys are overall less expensive than mail surveys (e.g., Illieva, Baron, and Healey 2002). Furthermore, the scale of online surveys is not associated with financial resources. Since the programming efforts are fixed and variable costs are virtually zero, large-scale surveys are as expensive as small-scale surveys given that the e-mail addresses are available. This democratizes the survey-taking process since it opens up new research opportunities for individuals and smaller corporations (Couper 2000). In addition, this scale effect facilitates cross-sectional studies and/or International comparisons (Illieva, Baron, and Healey 2002) since online surveys allow instant access to a wide audience, irrespective of geographical location. Another key advantage of online surveys is fast response. Illieva, Baron, and Healey (2002) conduct a literature review and find a mean response time of 5.59 days for online and 12.21 days for mail surveys. Moreover, online surveys make it easier to examine data as they are collected. Data entry errors can be avoided since no scanning forms or hand data entries have to be used by the researcher. Data are directly loaded in the data analysis software, thus saving time and resources associated with the data entry process (Illieva, Baron, and Healey 2002). Online surveys also allow for complicated designs through question branching and automatic routing, skipping patterns, forced answer prompts, personalization, and audio-visual material (Schillewaert and Meulemeester 2005).

Besides these advantages, several concerns about online marketing research exist. Couper (2000) for example discusses the effect of coverage, nonresponse and measurement error. According to him, coverage error, which is the difference between the target population and the sample frame, presents the biggest threat to the quality of online surveys. Even though the number of Internet users is constantly rising and starts to reflect the general population in some countries, still not everyone is online yet. Practitioners have often resorted to post-stratification or propensity score weighting to deal with coverage issues of online surveys. Yet, empirical evidence suggests that no simple weighting factor or adjustment strategy can make online samples representative (e.g., Robinson, Neustadt, and Kestnbaum 2002; Vehovar, Lozar Manfreda, and Batagelj 1999). It seems that online respondents are not only different with respect to demographics, but also with respect to other underlying but not yet identified factors. Nonresponse error refers to the problem that not every person from the sample is willing or able to participate in the survey (Couper 2000). Nonresponse error could be caused by interface problems where the respondent uses a different browser or other technical difficulties such as server problems. Furthermore, evidence indicates that online surveys suffer from lower response rates (Illieva, Baron, and Healey 2002), which raises the question whether there are significant differences between respondents and nonrespondents on the variables of interest. Since some surveys are completely anonymous and thus respondents' true identity is unknown, researchers might face the problem that they cannot even compare respondents on important background variables. Measurement error refers to the deviation between the information sought and information generated by the measurement used by the researcher. Similar to other self-administered survey modes such as mail surveys, measurement error in online surveys could arise from the respondents (lack of motivation, comprehension, deliberate distortion) or from the instrument (poor design, technical flaws) (Couper 2000). Yet, design plays a much larger role in online surveys where researchers have more tools available, but less control over what the respondent sees on his/her screen. Furthermore, there is little evidence on whether and how the designs of other survey modes, as for example established in the Tailored Design Method (Dillman 2000), translate to the Internet. Since a lot of online surveys nowadays are conducted via online panels, which bring the advantage that demographics are known and can be used in sampling, panel

conditioning can occur through ongoing participation in surveys and thus contribute to measurement error (Couper 2000). Also the issue of respondent motivation is crucial for online surveys where researchers are fighting for attention from the respondents. If invitations to participate in surveys are sent by e-mail, researchers compete with high-priority work-related e-mails and e-mails from friends and family. Banners or other forms of pure online surveys face other, more task-oriented or enjoyable demands from respondents' precious time.

Table 1.1 summarizes the main strength and weaknesses between online surveys and traditional ways of interviewing respondents.

TABLE 1.1: Differences between Online and Traditional Survey Modes

Criteria	Online Surveys	Face-2-Face	Mail Surveys	Telephone
Sample control	Low to moderate	Moderate to high	Low	Moderate to high
Control of Data Collection Environment	Low	Moderate to high	Low	Moderate
Control of Field Force	High	Low	High	Moderate
Quantity of Data	Moderate	Moderate to high	Moderate	Low
Use of Physical Stimuli	Moderate	Moderate to high	Low	Moderate
Perceived Anonymity	High	Low	High	Moderate
Social Desirability	Low	High	Moderate	Low
Obtaining Sensitive Information	High	Low	High	Moderate
Potential for Interviewer Bias	None	High	None	Moderate
Speed	Very high	Moderate	Low	High
Costs	Low	High	Low	Moderate

Adapted from: Malhotra, Naresh K. (1999), *Marketing Research: An Applied Orientation*, 3rd ed. New Jersey: Prentice Hall, p. 188

1.4.2 Types of Web Surveys

The various types of online surveys can be divided into nonprobability and probability approaches (Table 1.2). To clarify the types of online surveys that will be examined in this dissertation, the different forms will be briefly explained. For a detailed discussion see Couper (2000). *Polls as entertainment* are very popular but should not be confused with real surveys. Their main purpose is to entertain and exchange opinions (e.g., <http://www.misterpoll.com/>). *Unrestricted self-selected surveys* are one of the most prevalent forms of online surveys where respondents are recruited on portals, frequently visited Web Sites or dedicated survey sites. Since these surveys have no access restrictions, researchers have little control over multiple responses. *Volunteer opt-in panels* of Internet users have received most attention and present the fastest growing type of online survey. Prominent examples are Harris Poll Online, Greenfield Online, and Survey

Sampling International. These companies create volunteer panels by recruiting respondents on well-traveled sites and Internet portals. Demographic information is collected at the point of registration, which increases researchers' control of the sample selection for a particular survey. While quota or probability sampling methods are mostly used for inviting respondents to specific studies, a drawback of this type of web survey is the higher potential for self-selection of volunteers to join the panel.

Table 1.2 shows further that probability-based web surveys either restrict samples to populations with web access (Type 4 and 5), or alternative, traditional methods are used to identify and reach a broader sample of the population (Type 6, 7 and 8).

TABLE 1.2: Types of Web Surveys

Nonprobability Methods	Probability-Based Methods
Type 1: Polls as entertainment	Type 4: Intercept surveys
Type 2: Unrestricted self-selected surveys	Type 5: List-based samples
Type 3: Volunteer opt-in panels	Type 6: Web option in mixed-mode surveys
	Type 7: Pre-recruited panels of Internet users
	Type 8: Pre-recruited panels of full population

Source: Couper, Mick P. (2000), "Web Surveys: A Review of Issues and Approaches," *Public Opinion Quarterly*, 64 (4), 464-94.

In *intercept surveys*, every *n*th visitor of a website is asked to fill out a questionnaire. The sample is naturally limited to visitors of a specific website. The two main disadvantages are the optimal timing of the invitation interval and potential nonresponse error. *List-based samples* are available for populations with high coverage on the Internet. Examples are student lists or intra-organizational surveys. In *mixed-mode surveys*, online surveys are combined with other, traditional survey modes in order to increase response while minimizing costs. In contrast to opt-in volunteer panels (Type 3), *pre-recruited panels of Internet users* use probability sampling methods such as random digit dialing (RDD) to recruit respondents. The biggest threat is nonresponse, which can occur at several stages of the recruitment process, from the initial nonresponse to RDD to joining the panel and participating in online surveys. *Probability samples of full population* also use traditional probability approaches to elicit cooperation. Yet, in contrast to Type 7, researchers provide the necessary equipment for all respondents in exchange for their participation and

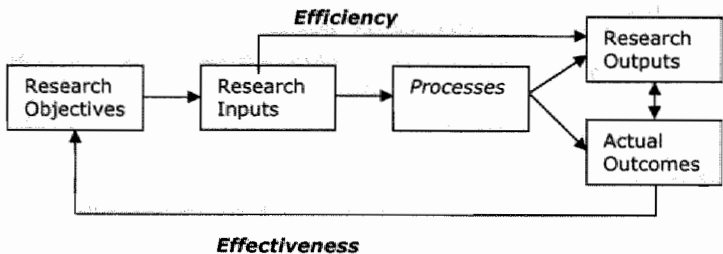
do not continue only with those respondents that already have access to the Internet. Major problems are low response rates and high costs.

The brief discussion of all types of Web Surveys shows that every method has its own advantages and disadvantages. All methods must be evaluated in the context of coverage and sampling error, nonresponse error, and measurement error (for a detailed discussion see Couper (2000) and Lozar Manfreda (2001)). The focus of this dissertation is on the most widely used and fastest growing volunteer opt-in panels and mixed mode surveys conducted with list-based samples.

1.5 Effectiveness of Online Marketing Research

The previous paragraphs demonstrate that the development of online marketing research is more than just a new research mode or a new way to gather information. Instead, online marketing research represents a technological and cultural change (Miller and Dickson 2001), as summarized succinctly in the introduction quote. The Internet provides instant information to millions of people worldwide and therefore has a profound impact on the way knowledge is created.

These technological advances have substantially broadened our research opportunities and offer tremendous advantages for surveyors. Online surveys are, for example, very efficient in terms of research costs and time. However, the crucial question is whether online surveys are also effective. As can be seen in Figure 1.1, effectiveness does not only require that research resources are used efficiently, but that the results also meet the initial research objectives. This implies that the data collected through online surveys is of adequate quality and meets the research objective.



Essentially, effectiveness of marketing research can be expressed as a function of two variables: (1) costs and (2) data quality. Putting these variables together, we can derive the following equation:

$$\text{Effectiveness} = w_0 + w_1 * \text{Costs} + w_2 * \text{Quality}$$

Where

Costs = all expenses related to conducting a research project (e.g. time, money)
Quality = quality of the data collected, appropriateness of the method to answer the research question

The weights of cost and quality are dependent on two main factors: (1) the magnitude of the risk of making a bad decision based on low-quality data, and (2) constraints such as limited research budgets or a need for fast results (Irons 2000). The equation provides the opportunity to allow for a great diversity of trade-off decisions and assists researchers in choosing the appropriate research mode. Yet, empirical research on online surveys is still in its infancy and few empirical studies have examined the effectiveness of online surveys. Researchers are left with questions such as: Can we trust the results of online surveys? Are data comparable to established research modes such as mail surveys? Is the quality influenced by advances in the design of online surveys that the online environment has enabled? How effective is the use of online access panels?

To answer these questions, this dissertation aims to enhance the online survey literature by examining the effectiveness of Internet-based marketing research. Effectiveness is a key element since it views research inputs (costs and time) and research outcomes (quality of data) in relation to research objectives. Thus, examining the effectiveness of online surveys lays the foundation for estimating the value of online surveys. The following paragraph will provide more details on the four empirical studies that attempt to shed more light on the effectiveness of online surveys.

1.6 Objectives of the Different Studies

The central research problem of this dissertation focuses on the effectiveness of online surveys. To provide comprehensive insights into several issues that are of importance for the evaluation of online surveys, four interrelated studies are conducted. Study one examines the influence of several design factors on response rate and response quality. The second and third

studies analyze the quality and generalizability of online surveys in comparison to mail surveys. Study four focuses on online access panels and examines what motivates respondents to join an online panel and whether different motives have an influence on response rate and response quality. The next section briefly describes all four studies and their respective objectives.

1.6.1 Objectives Study 1

Past research on traditional mail surveys has gathered extensive evidence on the influence of several format and design parameters, such as incentives or follow-ups, on response rate and response quality. Yet, the virtual environment of online surveys has added new facets to the discussion and administration of these factors so that results with regard to offline surveys are not per se generalizable to the online setting. Some authors have started to test the influence of certain format or design parameters in online surveys (Couper, Traugott, and Lamias 2001; Dillman et al. 1998; Lozar Manfreda, Batagelj, and Vehovar 2002; Sheehan and McMillan 1999). Yet, these studies are often narrow in scope and limit themselves to the assessment of response rates only. Therefore, the specific objective of this study is to examine the effect of the timing of follow-ups, different incentives, length, and presentation of the questionnaire on the response rate and response quality in an online experimental setting.

1.6.2 Objectives Study 2

Empirical evidence indicates that online surveys produce more socially desirable answers, more extreme responses, higher item completion, higher item variability, higher item omission, and higher measurement errors than mail surveys (Klassen and Jacobs 2001; Roster et al. 2004; Shermis and Lombard 1999; Stanton 1998). In contrast, other studies have found no significant differences between the online and traditional mail methods (Epstein et al. 2001; Knapp and Kirk 2003). Despite these findings, existing evidence regarding the comparison of Web-based and paper-and-pencil surveys remains both scarce and inconclusive. Assuming their equivalence still must be considered risky, because most of these studies have methodological limitations and conducted only limited statistical comparisons. Since comparable survey results are important,

especially for mixed-mode surveys, the objective of study two is to assess the equivalence of online and mail surveys.

1.6.3 Objectives Study 3

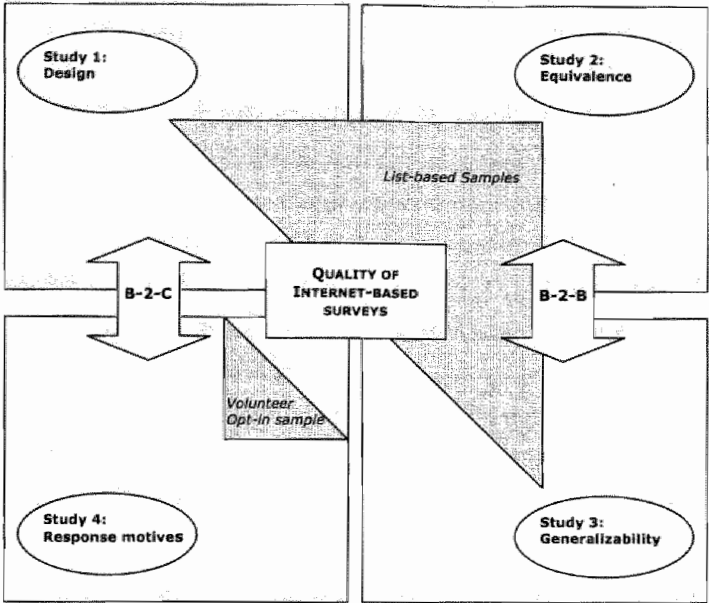
Several authors have found that online and mail surveys produce different results (e.g., McDonald and Adam 2003; Shermis and Lombard 1999) that cannot be equated through a simple weighting factor or adjustment strategy (e.g., Robinson, Neustadt, and Kestnbaum 2002; Vehovar, Lozar Manfreda, and Batagelj 1999). Yet, as mentioned earlier, evidence about the response quality of online surveys seems sparse and inconclusive (e.g., Epstein et al. 2001; Stanton 1998). In particular, virtually no survey research addresses the required or desired response quality in relation to data collection costs, despite the importance of the response quality/cost ratio in trade-off decisions. Costs and response quality are directly linked (e.g., if more items are added to a questionnaire to improve the reliability of the constructs and thus the response quality, the cost must also increase). Surveyors can thus pursue two different goals: (1) maximize response quality given a certain cost level or (2) minimize costs given the desired response quality level. Generalizability theory (G-theory) provides an interesting framework to evaluate the response quality of online surveys in relation to data collection costs. Therefore, the aim of study three is to address response quality and data collection costs of online and mail surveys simultaneously using G-theory.

1.6.4 Objectives Study 4

As mentioned in section 1.2.2, online access panels have received most attention and present the fastest growing type of online survey. Successful panels require a critical mass of members that are motivated to repeatedly participate in online surveys. In addition, uneven participation and panel attrition imply constant encouragement and recruitment of new respondents. Therefore, managers need to have a sound understanding of what motivates panel members. Yet, the primary focus of the survey literature has been on the effect of design factors and has neglected the importance of motivation. Therefore, the objective of study four is to develop a typology that systematically analyzes underlying individual differences that motivate respondents to participate. In addition, we examine whether different response motives affect response

parameters such as response rate and response quality, and whether respondents can be grouped into different clusters.

Figure 1.2: Integrative Framework



1.7 Outline of this Dissertation

This dissertation explores different issues with respect to Internet-based research. While the four studies investigate different theoretical aspects and employ different data gathering and analyses techniques, the four studies are clearly linked and build upon each other. Study one (chapter two) experimentally examines the impact of design factors on response rate and response quality in a consumer attitude and usage study. Once researchers know how to design online surveys, the crucial question about the response quality of web surveys emerges. Chapter three and four focus on this issue by comparing online and mail surveys in a B-2-B service quality survey. In study two (chapter 3), the equivalence between online and mail surveys is assessed, while study three (chapter four) uses a G-theory approach to assess the response quality of online surveys. Once systematic differences between online and mail surveys can be ruled out, researchers should focus on sampling issues for specific surveys.

Study four (chapter five), examines one of the most popular and fastest growing online survey techniques, namely volunteer opt-in online access panels. The specific focus lies on individual differences in motivation to participate in online surveys and the effect on response rate and response quality. Finally, chapter six presents a summary of our main results, discusses the theoretical and managerial implications as well limitations, and provides suggestions for future research. Table 1.3 offers a summary of the outline of this dissertation.

TABLE 1.3: Overview of the Dissertation

Chapter	Study	Research Methods
1	(Introduction)	
<i>Design of Online Surveys</i>		
2	Study 1: Response Rate and Response Quality of Internet-based Surveys: An Experimental Study	Experiments
<i>Comparability of Online and Mail Surveys</i>		
3	Study 2: An Assessment of Equivalence between Online and Mail Surveys in Service Research	Quasi-experiment
4	Study 3: Comparing the Generalizability of Online and Mail Surveys in Cross-National Service Quality Research	Generalizability theory
<i>Evaluating the Quality of Online Access Panels</i>		
5	Study 4: Individual Differences in Motivation to Participate in Online Surveys and the Effect on Response Rate and Response Quality	Surveys
6	(Conclusion)	

CHAPTER 2: RESPONSE RATE AND RESPONSE QUALITY OF INTERNET-BASED SURVEYS: AN EXPERIMENTAL STUDY²

This study examines the effect of the timing of follow-ups, different incentives, length, and presentation of the questionnaire on the response rate and response quality in an online experimental setting. The results show that short questionnaires have a higher response rate, although long questionnaires still generate a surprisingly high response. Furthermore, vouchers seem to be the most effective incentive in long questionnaires, while lotteries are more efficient in short surveys. A follow-up study revealed that lotteries with small prizes, but a higher chance of winning are most effective in increasing the response rate. Enhancing questionnaires with visual elements, such as product images, lead to a higher response quality and generate interesting interaction effects with the length of the questionnaire and the incentives used. Finally, the timing of the follow-up has no significant influence on the response rate.

² This chapter is largely based on Deutskens, Elisabeth C., Ko de Ruyter, Martin G.M. Wetzels, and Paul Oosterveld (2004), "Response Rate and Response Quality of Internet-based Surveys: An Experimental Study," *Marketing Letters*, 15 (1), 21-36.

*"With the graphic and multimedia capabilities of the computers, the survey researcher has an almost unlimited set of design choices in developing web-based surveys."
(Timmerman 2002)*

2.1 Introduction

The rapid growth of the Internet has opened new opportunities for collecting and disseminating research information worldwide. Market researchers have long recognized the advantages of Internet-based surveys, the most important of these being lower costs and faster response (e.g. Illieva, Baron, and Healey 2002). Existing research on online surveys has mainly focused on comparing response rates to traditional mail surveys or combining existing evidence into a meta-analysis of online survey response rates (e.g. Cook, Heath, and Thompson 2000; Illieva, Baron, and Healey 2002; Sheehan 2001; Shermis and Lombard 1999). Some authors have started to test the influence of certain format or design parameters (Couper, Traugott, and Lamias 2001; Dillman et al. 1998; Lozar Manfreda, Batagelj, and Vehovar 2002; Sheehan and McMillan 1999). These studies are often narrow in scope and limit themselves to the assessment of response rates only. There has been little methodological research around influences on response rates as well as response quality in online surveys. Before online surveys can be used on a large scale for academic as well as market research, it is necessary to experimentally determine the impact of format and design parameters on nonresponse error as well as on response accuracy and questionnaire completeness.

While empirical evidence on response rate and response quality of online surveys is still limited, past research on traditional mail surveys has gathered extensive evidence on the influence of several format and design parameters on response rate and response quality. The literature on mail surveys verifies that the most important factors for maximizing response rate and response quality are follow-up mailings and incentives. In addition to this, the length and presentation of the questionnaire have been examined extensively (Church 1993; Dillman 2000; Fox, Crask, and Kim 1988; Heberlein and Baumgartner 1978; Kanuk and Berenson 1975; Yammarino, Skinner, and Childers 1991; Yu and Cooper 1983). Yet, the virtual environment of online surveys has added new

aspects to the discussion and administration of these factors. We argue that results with regard to offline surveys are not per se generalizable to the online setting. On the one hand, online surveys share a lot of characteristics with traditional mail surveys (e.g. answering questions presented via text) (Kiesler and Sproull 1986). On the other hand, they differ in the method of contacting, the medium, and the mode of responding (Tourangeau, Rips, and Rasinski 2000). Thus, the interactive and impersonal nature of the Internet apparently causes differences in the effectiveness of format and design parameters between online and offline surveys. The aim of this study is to assist researchers in selecting the most effective incentives for online surveys, in choosing the appropriate length, and deciding upon the optimal timing of the follow-up contacts as well as on the optimum design of the online survey.

The remainder of the study is structured as follows: Firstly, in order to develop the hypotheses, empirical evidence from traditional mail surveys is combined with scarce empirical evidence from online surveys. Next, we describe the experimental study that we conducted, which aimed to test these hypotheses. A follow-up study provides some additional insights into the effectiveness of lotteries. Finally, we conclude this paper with a discussion and implications.

2.2 Literature Review and Hypotheses Development

Previous research has examined a large number of factors that increase response rates and improve data quality. In terms of online surveys, the most important factors to consider are follow-ups, incentives, and length and presentation of the questionnaire (Church 1993; Dillman 2000; Fox, Crask, and Kim 1988; Heberlein and Baumgartner 1978; Kanuk and Berenson 1975; Yammarino, Skinner, and Childers 1991; Yu and Cooper 1983). In the following paragraphs we will develop the hypotheses based on evidence from mail as well as online survey literature.

2.2.1 Follow-Ups

Follow-up contacts have been consistently reported as being the most powerful technique for increasing response rates, both in mail and online surveys (Dillman 2000; Fox, Crask, and Kim 1988; Heberlein and Baumgartner 1978; Schaefer and Dillman 1998; Yammarino, Skinner, and Childers 1991). Dillman

(2000) suggests the use of four contacts with a participant, but even single follow-ups have been reported to increase the response rate significantly (Heberlein and Baumgartner 1978). Sending out multiple follow-ups in an online survey is virtually costless, however, it should be done with great care. Repeated follow-ups have diminishing returns and may be considered as spam, thereby irritating or annoying potential respondents without noticeably increasing response rates (Solomon 2001).

Sending the follow-up right after the majority of respondents have reacted to the initial mailing has been identified as essential for maximizing the response rate (Dillman 2000). At the same time, the question of the exact timing between the initial and subsequent request has been left largely unresolved. Essentially, the response rate takes on the form of an inverted U-shape. However, the central question is which distribution characteristics the response curve for online surveys adopts, as this determines the best possible timing of the follow-up. The literature review from Illieva, Baron and Healey (2002) reveals that the average response time in online surveys is 5.59 days, which is substantially faster than the 12.21 days in mail surveys. Hence, follow-ups could be conducted earlier in online settings than in traditional settings, for example after one week (Dillman 2000). As the optimal timing of the follow-up is imperative for maximizing the response rate, we hypothesize that a follow-up, which is sent right after the majority has responded to the initial mailing, will increase the overall response rate in an online survey compared to a later follow-up.

With respect to response quality, neither research on mail surveys nor experiential or intuitive reasoning suggests an effect of the timing of follow-ups on response quality. Therefore, we do not expect, a priori, a direct effect of the timing of follow-ups on the response quality of an online survey. Hence we hypothesize:

H_{1a}: *Earlier follow-ups will increase response rates compared to later follow-ups.*

H_{1b}: *The timing of the follow-up will not influence response quality.*

2.2.2 Incentives

In addition, the use of monetary incentives in general, and small prepaid financial incentives in particular, have been declared as being effective in increasing the response rate in offline and online surveys (Church 1993; Dillman 2000; Fox, Crask, and Kim 1988; Heberlein and Baumgartner 1978; Kanuk and

Berenson 1975; Yammarino, Skinner, and Childers 1991; Yu and Cooper 1983). However, the intangible nature of the Internet has raised new (administrative) problems, as cash incentives cannot be attached to a virtual questionnaire. Therefore, empirical research is needed to identify alternative online incentive systems, such as vouchers, lotteries or donations, and examine their effect on response rates. As empirical evidence from mail surveys verifies that cash is more effective than lotteries or charitable donations (Furse and Stewart 1982; Warriner et al. 1996), we hypothesize that the response rate in an online survey will be highest if vouchers, the closest online equivalent to cash, are used as an incentive, followed by lotteries and finally donations.

Self-perception theory of respondent behavior stipulates that if external cues, such as incentives, are present respondents feel less committed and hence provide lower quality responses (Hansen 1980). However, the mail literature does not provide evidence that the response quality varies for specific types of incentives. As there is no evidence (or a logical reason) to assume that any of the three incentives will affect response quality, we hypothesize that the response quality is the same regardless of whether vouchers, lotteries, or donations are used as an incentive. Therefore, we postulate:

H_{2a}: *Response rates will be highest if vouchers are used as an incentive, followed by a lottery and finally donations.*

H_{2b}: *Response quality will be equal for vouchers, lotteries or donations.*

2.2.3 Length

The relationship between length and response rate and quality may be different for Internet-based research, as it is not clear what 'long' means in the online environment. Rosenblum (2001), for example, indicates that online surveys should consist of approximately 20 questions, which would generally be considered too short for substantial market and academic research. Therefore, we need to investigate whether response rate and response quality are indeed lower for longer online surveys. In terms of mail surveys, the length of the questionnaire has been repeatedly investigated (Dillman 2000; Fox, Crask, and Kim 1988; Heberlein and Baumgartner 1978; Kanuk and Berenson 1975; Yammarino, Skinner, and Childers 1991; Yu and Cooper 1983). Common sense suggests that longer questionnaires will obtain lower response rates than shorter questionnaires, as they demand more time from the respondent. Likewise, one

would also assume a negative relation between the response quality and length of the questionnaire. However, the literature about traditional mail surveys provides mixed results. Whereas several studies show that survey length does not influence response (Linsky 1975; Yu and Cooper 1983), a number of studies reveal that there is a negative relation between survey length and response rate as well as response quality (c.f. Heberlein and Baumgartner 1978; Yammarino, Skinner, and Childers 1991). Therefore, we hypothesize:

H_{3a}: *Short questionnaires will increase response rates compared to long questionnaires.*

H_{3b}: *Short questionnaires will exhibit a higher response quality compared to long questionnaires.*

2.2.4 Presentation of the Questionnaire

Previous research on paper-and-pencil surveys suggests that the design of the questionnaire may be extremely important in obtaining unbiased answers from respondents, as respondents evaluate both the verbal and the visual elements of the questionnaire. Dillman (2000) states that "respondent-friendly" design improves mail survey response, but that there is a lack of agreement on what exactly constitutes respondent friendly design. Research on specific graphical aspects, such as color of the questionnaire, does not provide clear results about their effect on response rate and response quality (Fox, Crask, and Kim 1988; Kanuk and Berenson 1975).

The Internet has added a new dimension to the design of questionnaires as it offers a wide area of new design opportunities. Simple questionnaires can, for example, be enhanced with cinematic and interactive images, such as 3-D presentations of products. This results in web questionnaire design challenges for the researcher. On the one hand, pictures may enhance the attractiveness of the survey and may make it a more enjoyable experience for the respondent. In this event, it is likely that the response quality would be higher for surveys that are graphically enhanced, as respondents enjoy the process of filling in the questionnaire and therefore put in more effort and answer more seriously. On the other hand, these advanced features make the questionnaire more difficult to access and complete and lead to longer download times, which could consequently reduce the response rate (Dillman 2000). Therefore, we hypothesize:

H_{4a}: *Visually enhanced questionnaires will decrease response rates compared to text-based questionnaires.*

H_{4b}: *Visually enhanced questionnaires will improve response quality compared to text-based questionnaires.*

2.3 An Empirical Study

2.3.1 Research Setting

The questionnaire used in this research was a standardized, multi-client attitude and usage study, which has been conducted in the Netherlands via a traditional mail surveys since 1991. The questionnaire contained 19 product categories (such as beer, shampoo, cheese, olive oil and toast) in the long version and 9 product categories in the short version of the survey. Every product category included a number of brands for which respondents had to indicate whether they are familiar with the brand, how often they use it and how they perceive the price and quality of this brand. Participants in this experiment were recruited from a database, which consists of names and e-mail addresses from people who had participated in a prior telephone survey and consented to being contacted again for other research purposes. Respondents received an e-mail invitation for this research, which included a short introduction to the study, a request to participate and the hyperlink to the web questionnaire. With one click on this link, respondents were directed towards the questionnaire. By using a unique ID for each respondent, respondents did not have to fill out a username and password before entering the questionnaire.

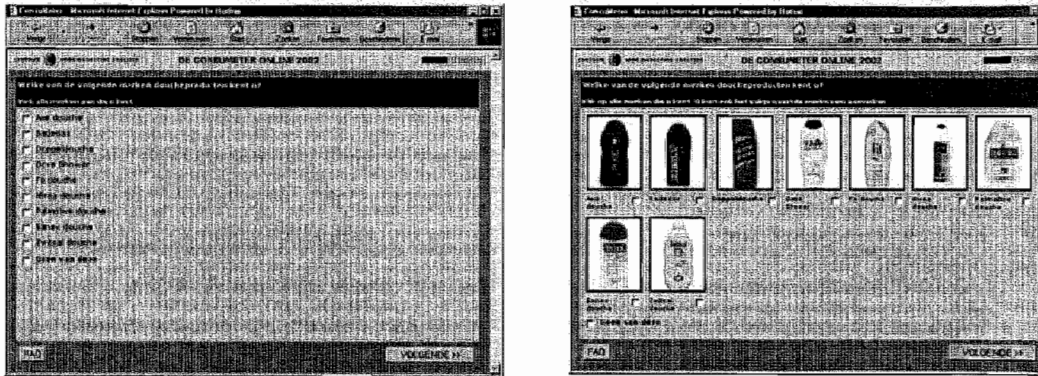
2.3.2 Independent Variables

To test the hypotheses, we designed a between-subjects, fixed effects factorial design with four factors. These were: type of incentive (voucher, lottery, donation), length of the questionnaire (long, short), presentation of the questionnaire (textual, visual), and timing of the follow-up (early, late). This resulted in a 3 x 2 x 2 x 2 full factorial design with 24 different cells.

First, we used three different incentives in this experiment: *Vouchers*: a 2 Euro voucher in the short version, or a 5 Euro voucher in the long version for an online book and CD store. We opted for an online book and CD store, as "books and CD's continue to be the most popular items bought online for the 3rd consecutive year" (Greenspan 2002). *Charitable donations*: we noted the overall

amount that we would donate if everybody in the survey participated (€500). Respondents could choose between the World Wide Fund for Nature (WWF), Amnesty International, or a cancer association. *Lottery*: respondents had the chance of winning one of 5 vouchers of 25 Euro in the short version and 50 Euro in the long version for an online book and CD store. The second independent variable was *length* of the questionnaire, where the long version took approximately 30-45 minutes to finish, while the short version could be completed within 15 to 30 minutes. Thirdly, the presentation of the questionnaire was either *textual* or *visual*. To utilize the design possibilities on the Internet, we presented one random half of the respondents with only the name of the product, while the other half of the respondents obtained in addition to the brand name also a picture or logo of the product. Finally, we randomly divided participants, who did not fill-in the questionnaire after the initial invitation, into two groups, an *early* and a *late* follow-up group. Based on initial empirical evidence that the mean response time to online surveys is 5.59 days (Illieva, Baron, and Healey 2002), the 'early' group received the reminder after one week. The 'late' group received the follow-up after two weeks, as suggested by the mail literature (Dillman 2000).

Figure 2.1: Textual and Visual Version of the Questionnaire



2.3.3 Dependent Variables

The two dependent variables that were analyzed in this study were response rate and response quality. Response rate was defined as the percentage of the contacted sample that has answered and returned the questionnaire. We only

considered the net response rate, i.e. the percentage of questionnaires that actually reached the respondent, thus excluding undeliverable e-mails.

With respect to response quality, two specific components of data quality were examined, namely the completeness and accuracy of respondent answers (Goetz, Tyler, and Lomax Cook 1984; Hansen 1980; McDaniel and Rao 1980). Since no open-ended questions were used in this survey, completeness was assessed by the number of non-opinions ("don't know" answers) given by each respondent (Goetz, Tyler, and Lomax Cook 1984) and the number of semi-completed questionnaires. "Don't know" answers were asked after respondents had indicated that they know a product, but were not able or willing to answer the price, quality and usage questions with respect to that product. Thus, "don't know" answers were considered undesirable, as respondents were either inconsistent with previous answers, or they wanted to speed up the process by indicating "don't know". Semi-completed questionnaires refer to questionnaires where respondents started to fill them in, but did not complete them.

The second aspect of response quality was bias or inaccuracy (Goetz, Tyler, and Lomax Cook 1984). As early as 1978, Hansen and Scott noted that "bias" is an umbrella term that covers a broad range of quality problems. Some authors look at bias as a deviation from a "known" and presumably truthful response" (Hansen 1980, p. 79). Yet, Hansen (1980) suggested in his JMR study, where he identified a negative link between incentives and response quality, that the distribution or summary of responses of one subgroup should be compared with those of another subgroup. Therefore, we examined the response distribution (Hansen 1980) by comparing means and variances of the answers, as this method is especially appropriate "when different methods are used in an attempt to stimulate response to a survey" (p. 79).

2.4 Results

Data collection took place from mid April until the end of May 2002. The profile of the respondents shows that there is a balance in gender, with 51.5% male and 48.5% female respondents. Younger and higher educated people are somewhat over-represented. For example, 36.2% are between 15-34 years old, 40.0% are in the range of 35-49 years, still 22.1% are 50-64 years, but only 1.8% of the respondents are 65-74 years old. With respect to education, 50.4% completed higher education, 42.6% got a medium level of education, and only

5% lower education. After only three days, we had received more than half of the final responses (52.9%). On average, it took respondents 6.6 days to complete and return the questionnaire, which is slightly higher than the 5.59 days reported by Illieva, Baron and Healey (2002). Interestingly, we observed that respondents in the lottery group responded with 5.7 days significantly faster than the donation group with 6.7 days and the voucher group with 7.4 days ($F(2,727) = 3.81, p = 0.023$). In addition, we employed the Games-Howell procedure for multiple comparisons of the response time (Games and Howell 1976) and found that the voucher and lottery group differed significantly ($t = 2.84, p = 0.013$).

In total, 5,413 e-mails were sent out. One third of those e-mails (1,836) were undeliverable, which left us with 3,577 usable e-mails, from which we received 730 completed questionnaires. This yielded a net response rate of 20.4%. The response rate in the online survey differed significantly between the different cells of the design ($\chi^2(23) = 73.97, p < 0.001$; $G^2(23) = 77.24, p < 0.001$). The short, visual version of the questionnaire with lottery as incentive and a late reminder had the highest response rate with 31.4%, while the long, visual version with donation to charity as incentive and an early reminder only had a response of 9.4%. Table 2.1 summarizes the response rates for the main effects.

TABLE 2.1: Response Rates for the Different Format and Design Parameters

Factors						
Type of Incentive	Length of the Questionnaire		Presentation of the Questionnaire		Timing of the Follow-up	
Voucher	22.8%	Long	17.1%	Textual	21.9%	Early
Donation	16.6%	Short	24.5%	Visual	19.0%	Late
Lottery	22.8%					

Testing Hypotheses

$\chi^2(2) = 20.26 (p < 0.001)^a$	$\chi^2(1) = 29.94 (p < 0.001)$	$\chi^2(1) = 4.55 (p = 0.033)$	$\chi^2(1) = 1.70 (p = 0.192)$
$G^2(2) = 20.63 (p < 0.001)^b$	$G^2(1) = 29.72 (p < 0.001)$	$G^2(1) = 4.54 (p = 0.033)$	$G^2(1) = 1.70 (p = 0.192)$

a Pearson Chi-Squared Statistic

b Likelihood-Ratio Statistic

In order to rule out non-response error, we carried out a time trend extrapolation test. The assumption is that respondents who respond less readily are similar to non-respondents (Armstrong & Overton, 1977). The early respondents included all those who filled in the questionnaire on the first day

after they had received the e-mail invitation, whereas the late-respondent group consisted of the people who responded after 10 days or later. Only a negligible number of variables (less than 1%) used in the questionnaire showed a significant difference between early and late respondents. Those variables were distributed among all items of the questionnaire so that no consistent pattern could be discerned. Therefore, we may conclude that our data did not suffer from nonresponse problems.

2.4.1 Follow-Ups

In line with our expectations, the early follow-up had a higher response rate (21.2%) than the late follow-up (19.5%). However, this difference was not significant ($\chi^2(1) = 1.70$, $p = 0.192$; $G^2(1) = 1.70$, $p = 0.192$) and hence, H_{1a} was not supported. As hypothesized, we did not find a significant difference in the response quality for the early and the late follow-up. Neither the analysis for completeness, nor for accuracy, revealed significant differences at $\alpha = 0.05$. Thus, H_{1b} was supported.

2.4.2 Incentives

Vouchers and lotteries with 22.8% had a significantly different response rate ($\chi^2(2) = 20.26$, $p < 0.001$; $G^2(2) = 20.63$, $p < 0.001$) to the donation to a charity group with 16.6% (where 61% choose for the cancer association, 25% for the WWF, and 15% for Amnesty International, adjusted residual = $|4.50|$). Thus, H_{2a} was only partially supported. As hypothesized, donations had a lower response rate than both vouchers and lotteries. However, the overall response rate of vouchers was not higher than the response rate of lotteries. ANOVA revealed that there were no significant differences in the number of "don't know" answers at $\alpha = 0.05$, and also the number of semi-completed questions did not vary between the different incentive groups ($\chi^2(2) = 2.69$, $p = 0.261$; $G^2(2) = 2.67$, $p = 0.264$). In addition, the analysis of accuracy did not show significant differences at $\alpha = 0.05$. As we did not find a discrepancy in the response quality for the different types of incentives, H_{2b} was supported.

2.4.3 Length

As expected, the short version of the questionnaire had a significantly higher response rate ($\chi^2(1) = 29.94$, $p < 0.001$; $G^2(1) = 29.72$, $p < 0.001$) than

the long version of the questionnaire, with 24.5% and 17.1% respectively. Hence, H_{3a} was supported. The analysis of the number of "don't knows" in the long and the short version revealed that there are proportionally more "don't know" answers in the long version at $\alpha = 0.05$. Also the analysis of the number of semi-completed questionnaires showed that there are more semi-completes in the long version than in the short version (220 semi-completed questionnaires in the long version versus 131 in the short version; $\chi^2(1) = 22.58, p < 0.001$; $G^2(1) = 22.79, p < 0.001$). Furthermore, respondents stopped relatively earlier in the long version (after 41.56%) than in the short version (54.29%) of the questionnaire ($p < 0.001$). For the second aspect of response quality, bias or inaccuracy, ANOVA did not reveal a significant difference in means and variances at $\alpha = 0.05$ for the length of the questionnaire. Aggregating the findings from the completeness and accuracy analysis, we found partial support for H_{3b} .

2.4.4 Presentation of the Questionnaire

The visual presentation of the questionnaire had, with 19.0%, a significantly lower response ($\chi^2(1) = 4.55, p = 0.033$; $G^2(1) = 4.54, p = 0.033$) than the textual presentation with 21.9% and therefore H_{4a} was supported. In terms of presentation of the questionnaire, 5 out of the 19 product categories produced a significant difference at $\alpha = 0.05$ in the number of "don't know" answers. Respondents in the textual group indicated more often "don't know" than respondents in the visual group. In the analysis of the number of semi-completed questionnaires in the visual and textual version of the questionnaire, we did not find a significant difference ($\chi^2(1) = 0.551, p = 0.458$; $G^2(1) = 0.551, p = 0.458$), although participants in the long version of the questionnaire stopped earlier in the visual version (45.39% answered questions versus 52.13%), whereas respondents in the short version of the questionnaire stopped earlier in the textual version (56.95% versus 60.63%). Comparing the means and variances for the visual and textual version did not show a significant difference at $\alpha = 0.05$. Hence, the evidence for H_{4b} was equivocal. Respondents in the textual version indicated more often "don't know", but there were no significant differences in the means and variances and the analysis of the semi-completes did not reveal a main effect for the presentation of the questionnaire.

2.4.5 Interaction Effects

In the previous sections we used contingency table analysis for testing the main effects of the factors in this study (Everitt 1992). To further refine our results, we also tested for interaction effects of incentives, timing of follow-ups, length and presentation of the questionnaire, on response rate and response quality. For testing the impact on the response rate, we employed logit modeling (Agresti 1990; Everitt 1992). We used the HILOGLINEAR procedure in SPSS release 11.0 using a backward elimination approach. The partial association tests indicate that two two-factor interactions are significant at $\alpha = 0.05$, i.e. INCENTIVE*LENGTH and INCENTIVE*PRESENTATION. In the short version, lotteries were most effective (27.9%), while vouchers and donations had with 23% the exact same response rate. In the long version, vouchers performed best (22.6%), followed by lotteries (18.8%) and donations (12.7%). With respect to the presentation of the questionnaire, we see that in the textual version of the questionnaire, the response rates are about the same (lottery: 22.9%, voucher 22.4% and donations 20.5%), while the response rate for the donation group drops to 14% in the visual version. In addition, we used the LOGLINEAR procedure implemented in SPSS release 11.0 to conduct a hierarchical logit analysis using nested models. The results suggested that omitting the two-factor interaction terms from the model significantly decreased the fit of the model ($\chi^2(18) = 26.34$, $p = 0.092$; $G^2(18) = 26.61$, $p = 0.087$; $\Delta\chi^2(9) = 21.47$, $p = 0.011$, $\Delta G^2(9) = 21.74$, $p = 0.009$). Inspection of the z statistic for the individual parameter estimates confirmed that the two-factor interaction terms (INCENTIVE*LENGTH) and (INCENTIVE*PRESENTATION) were significant at $\alpha=0.05$. The final model retained the "main" effects (INCENTIVE) and (LENGTH) and the two-factor interaction terms (INCENTIVE*LENGTH) and (INCENTIVE*PRESENTATION). The results suggest a good fit to the data ($\chi^2(16) = 10.92$, $p = 0.814$; $G^2(16) = 10.94$, $p = 0.813$).

The interaction effects of response quality were analyzed with the GLM Univariate procedure to carry out a series of ANOVAs. However, the analysis did not reveal significant interaction effects for the means and variances at $\alpha = 0.05$. Also the number of "don't knows" and semi-completes were not significantly different at $\alpha = 0.05$ between the different treatment groups.

2.4.6 Additional Study on Incentives

The results indicate that lotteries present a very interesting new form of incentives. Participants in the lottery group of the first experiment had the chance of winning one of 5 vouchers of 25 Euro in the short version and 50 Euro in the long version of the questionnaire for an online book and CD store. When selecting the prize type and structure, we based our choice on evidence from the mail literature as well as best practice from a large marketing research agency. Yet, it could also be examined whether an attractive commodity, such as a DVD player, as a first prize in the lottery would be preferred to an online voucher. To further explore whether we used the most effective lottery type, we conducted a follow-up study where we focused on one of our most interesting findings. In total, 497 students were contacted via e-mail and asked to participate in an online survey on the perceived quality of examination facilities at a large university. The questionnaire contained 4 open-ended questions and 40 closed questions, which were evaluated on a 7-point Likert scale ranging from "totally disagree" (1) to "totally agree" (7). Respondents were randomly assigned to one of the lotteries, where they were informed that they had the chance of winning either one out of ten vouchers of 25€, one out of five vouchers of 50€ or one DVD player. By focusing on lotteries only, we manipulated the values of the lotteries and separated it from the length of the questionnaire. Hence, we could ensure that the overall expected monetary value between all treatment groups was the same.

The overall response rate was 28.2%. The final sample consisted of 42.7% female and 57.3% male respondents. With respect to the effectiveness of the different types of lotteries, the findings indicated that the lottery with 10*25€ had a significantly higher response rate (35.5%) than the lottery with 5*50€ (26%) and the DVD player (22.5%) ($\chi^2(2) = 7.43$, $p = 0.024$; $G^2(2) = 7.35$, $p = 0.025$). In terms of response quality, the means and variances of the closed questions were analyzed, as well as the number of words and arguments that respondents used in their answers to open-ended questions. No significant differences at $\alpha = 0.05$ could be identified for the variances of the closed questions. However, the means were significantly different for three questions, where two times the mean of the DVD-player as incentive was significantly lower than the mean of the 25€ version ($\alpha = 0.050$ and 0.035 respectively), and once the DVD was significantly higher than the 25€ and 50€ versions ($\alpha = 0.005$). No

significant differences were found in the variances of the number of arguments and the variances and means of the number of words. However, for two open-ended questions, the mean number of arguments differed significantly ($\alpha = 0.03$ and $\alpha = 0.02$ respectively), where the mean was lowest for the DVD player as an incentive. Thus, lotteries with smaller prizes but a higher chance of winning are most effective in increasing the response rate and are also favorable in terms of response quality.

2.5 Discussion and Implications

This study examines the effect of several design parameters on response rate and response quality in online surveys. With respect to the optimal timing for sending the follow-up, the results do not give a clear answer. There seems to be a slight preference for sending the follow-up after one week instead of after two weeks, although the results were not significant. Nevertheless, we propose that reminders are sent quite early in order to fully utilize the fast turnaround times of online questionnaires in comparison to mail surveys. The results indicate, in line with previous research, that monetary interests (vouchers and lotteries) are stronger than altruistic appeals (donations) in increasing the response rate of online surveys. This effect is especially important for long questionnaires. Respondents in the long version demanded an incentive that seemed to reimburse them for the time and effort they devoted to filling in the questionnaire, as can be seen in the remarkably low response rate of 12.7% for the donation group. Furthermore, respondents in the short version appeared to be more risk-taking, as lotteries were by far the most effective type of incentive. Interestingly, vouchers and donations received the exact same response rate in the short version, which could be due to the fact that respondents in the short version had to devote less time and effort and were hence more willing to act altruistically and accept incentives that did not compensate themselves. Overall, we propose that lotteries are probably the most effective reward in an online environment, as they lead to the highest response rate in the short version and still a respectable response in the long version, while being much more cost-efficient than vouchers. In addition, the response time in the lottery group was more than 1.5 days faster than in the voucher group, as respondents in the lottery might have inferred that by responding early, they have a higher chance of winning a prize. To explore the effectiveness of lotteries further, we conducted

a follow-up study that examined whether an attractive commodity such as a DVD-player would be preferred to different levels of online vouchers. In line with our prior reasoning, we found a significantly lower response rate for the DVD player and a slight difference in response quality. Hence, a lottery with smaller prizes but a higher chance of winning is most effective. The effectiveness of using a commodity such as a DVD player is likely to depend on the attractiveness as well as usefulness of that prize to consumers.

As hypothesized, the short version of the questionnaire had a higher response rate than the long version. However, the myth that online surveys should not contain more than 20 questions (Rosenblum 2001) may be corrected, as even the short version with a response time between 15-30 minutes was still relatively long for an online survey. Hence, response rates of almost 25% in the short version and still around 17% in the long version illustrate that even relatively long questionnaires yield a considerable response rate. Most importantly, the length of the questionnaire did not have a negative effect on the quality of responses. Only the number of "don't knows" and semi-completes were slightly higher in the long, and, in particular, in the long, visual version of the questionnaire. Thus, it seems feasible to conduct long and elaborate surveys via the Internet, especially when respondents are adequately rewarded.

Examining the main effects of the presentation of the questionnaire on the response rate, we find that the response rate is significantly lower for the visual than for the textual version of the questionnaire, although this difference is relatively small. However, when incorporating interaction effects, the main effect becomes insignificant and hence the interaction effects are much more meaningful. The analysis of the number of semi-completes revealed that participants in the long version of the questionnaire abandoned the questionnaire significantly earlier in the visual version, while respondents in the short version of the questionnaire stopped earlier in the textual version. This may suggest that respondents in the long and visual version of the questionnaire were faced with long downloading times due to the large number of pictures, which in turn had a negative impact on their motivation to continue with the questionnaire and hence on their response rate. For the short and visual version on the other hand, the download time did not seem to be problematic, as the short-visual-lottery condition had the highest overall response rate. In any case, as computer and Internet connection speed increases in the future, the negative effect of longer

downloading times of visual presentations will diminish over time. Interestingly, the effectiveness of the different types of incentives differed largely between the respondents in the textual and visual version. Whereas all incentives performed almost equally well in the textual version, the results indicated that in the visual version, the response rate for the donation group dropped considerably from 20.5% to only 14%. This supports our assumption that respondents in the visual version were subject to longer downloading times, and therefore demanded an incentive that compensated for the time and effort they spent on the questionnaire. However, respondents who had completed the visual version of the questionnaire answered less often with "don't know" than respondents in the textual version. It seems that once respondents recognize a product on the basis of the name and the picture, they have a more active association of the relevant features of the product and hence are better able to give their opinion about the price and quality of the product.

2.6 Limitations and Future Research Guidelines

Future research on Internet-based surveys should be directed at confirming the effects of the timing of follow-ups, incentives, presentation and length of the questionnaire across different questionnaires and populations. The studies were conducted in the Netherlands, which is especially suitable for this type of research in Europe, since according to Nielsen-Netratings (NUA Internet Surveys 2003) it is rated one of the most mature Internet markets (together with Sweden, Hong Kong, and Australia). Furthermore, the Netherlands is the fourth country in terms of Internet connection speed on home PC (Hupprich and Bumatay 2002), outperforming countries such as the US and UK. Hence, the findings from this article serve as a good signal of how incentives, visual presentation, the timing of the follow-up, and length of the questionnaire will influence the response rate and response quality in countries where the Internet penetration is still lower.

For this study, we had access to a large database, which contained e-mail addresses from respondents to an earlier telephone survey. Further research should test whether the findings hold if a sample is recruited differently, for example with the help of an online panel, through site intercepts or pop-ups, or by using a multimode strategy. In this study, we opted to use incentives in all cells of the experiment, as we felt that the length and complexity of the questionnaire necessitated some form of compensation for the respondents,

especially as the research was very general and participants did not associate themselves with the topic of the questionnaire. Hence, the experiment should be repeated with a no-incentive group in a context where respondents hold a relation to the research, as for example in a customer satisfaction survey, where incentives are presumably not necessary and could lead to a response rate that is only marginally lower than for the incentive group. In this experiment, respondents in the lottery group were not notified about the number of participants in the lottery so that they were unable to calculate the expected value of the prize. Given that online surveys are usually sent to many people, respondents in the lottery condition might have perceived their chances of winning, and thus the expected value of their prize, as much lower. Future studies should ensure that respondents receive sufficient information to calculate the expected value and whether it is equal across the different incentive groups.

Caution should be exercised in applying the results of this paper to other types of studies where respondents might have different motivations for responding, for example surveys about the attitudes, opinions, and lifestyles of online respondents. In addition, the insignificant findings with respect to the timing of the follow-ups could indicate that neither the early follow-up after one week, nor the late follow-up after two weeks were ideal. Therefore, we suggest that future studies focus on this aspect and re-examine the optimal timing of follow-ups in online surveys. Insights into these issues will advance the knowledge base about online surveys and thereby help to empirically assess the potential of the Internet to supplement traditional means of conducting research by ensuring an adequate response rate and response quality.

CHAPTER 3: AN ASSESSMENT OF EQUIVALENCE BETWEEN ONLINE AND MAIL SURVEYS IN SERVICE RESEARCH³

This study examines whether online and mail surveys produce convergent results, which would allow them to be used in mixed-mode service quality studies. In the context of a large B2B service quality assessment, an analysis of the accuracy and completeness of respondent answers to both open and closed questions suggests that online and mail surveys produce equivalent results. Both the composite reliability and the average variance extracted show consistently high levels for both groups, and the means and variance-covariance matrices are equal across modes. However, minor differences occur between the two survey methods; online respondents provide more improvement suggestions, indicate more often to which competitor they want to switch, and provide lengthier answers in response to requests for examples of positive experiences with the company. This research thus provides important findings regarding the process for and results of comparing two survey modes.

³ This chapter is largely based on Deutskens, Elisabeth C., Ko de Ruyter, and Martin G.M. Wetzels (2006), "An Assessment of Equivalence between Online and Mail Surveys in Service Research," *Journal of Service Research*, Forthcoming.

*"Is the information that you retrieve from the Internet accurate, reliable, and valid?"
(Forrest 1999, p. 11)*

3.1 Introduction

Both academics and practitioners identify service excellence as a key factor in today's business environment (e.g., Zeithaml and Bitner 2003). To evaluate their services, most companies turn to their customers and collect customer-perceived service quality data on an ongoing basis, but the continuous measurement of service quality is both costly and time consuming. As a result, firms increasingly make use of online surveys to collect data about service performance.

Online surveys offer great advantages over traditional mail surveys, such as lower costs and faster responses (e.g., Illieva, Baron, and Healey 2002; Schuldts and Totten 1994). However, critics have questioned the quality of responses gathered through online surveys and suggested that the completeness and accuracy of the data they provide may not match those of traditional mail surveys.

In addition, because different survey modes often produce different results (Dillman 2000), comparable survey results are especially important for mixed-mode surveys, in which companies use both online and mail surveys in combination to reduce both their costs and nonresponse rates. Evidence from previous research has verified that online and telephone surveys may exhibit different underlying factor structures (Roster et al. 2004). Other studies indicate that online surveys produce more socially desirable answers, more extreme responses, higher item completion, higher item variability, and higher measurement errors (Klassen and Jacobs 2001; Shermis and Lombard 1999; Stanton 1998). In contrast, still other studies have found no significant differences between the online and mail methods (Epstein et al. 2001; Knapp and Kirk 2003).

Despite these findings, existing evidence regarding the comparison of Web-based and paper-and-pencil surveys remains both scarce and inconclusive. Assuming their equivalence still must be considered risky, because most of these

studies have methodological limitations and conducted only limited statistical comparisons.

This study contributes to the extant service research literature in several important ways. Due to the inevitable time and resource constraints involved in service quality survey research, online surveys represent attractive alternatives or supplements to traditional mail surveys. However, mixed-mode surveys can serve as a trustworthy basis for decision making only if convergent validity between the two survey methods can be established. Therefore, we empirically examine the response quality and measurement invariance of Internet-based and traditional mail surveys in a service quality study. In addition, we attempt to align the diverse literature on the comparison between online and offline surveys by indicating how different aspects of response completeness and accuracy might be analyzed in the future. Furthermore, because the existing literature on measuring the response quality of surveys is sparse, we use a service quality survey from a large office equipment manufacturer to show how the measurement invariance (Jöreskog 1971; Vandenberg and Lance 2000) and response quality (Goetz, Tyler, and Lomax Cook 1984; Hansen 1980; McDaniel and Rao 1980) of online and mail surveys can be analyzed and compared. Also, we provide managerial guidelines on the combined use of both online and mail surveys that will enable practitioners to optimize their response rates, response times, and survey costs while maintaining high response quality. Finally, this study addresses the critical issue of obtaining high-quality, representative online samples in service research.

The remainder of the study is structured as follows: we begin with a review of the existing empirical evidence on the equivalence of online and mail surveys. Next, we assess the equivalence of on- and offline surveys in a service quality context. Finally, we conclude with a discussion of the results and their theoretical and practical implications.

3.2 Literature Review

Mixed-mode surveys combine two survey modes, such as online and mail surveys, to compensate for the weaknesses of each. For example, various persuasive arguments recommend online surveys over traditional methods because of their lower costs, faster response, and wider geographic reach (e.g., Green, Johnson, and Neal 2003; Illieva, Baron, and Healey 2002; Schuldt and

Totten 1994). Furthermore, proponents of online surveys argue that the Internet provides uncomplicated directions (e.g., automatic routing), as well as richer and more interesting question formats (Klassen and Jacobs 2001; Simsek and Veiga 2001). However, in terms of coverage, the potential for invalid addresses, and representativeness, mail surveys are still preferred. Table 3.1 provides an overview of the strengths and weaknesses of both methods.

A mixed-mode design thus can reduce overall costs while maximizing response rates and minimizing nonresponse. But if online and mail surveys produce different results for the same study, their data cannot be aggregated. In Table 3.2, we provide an overview of existing empirical evidence regarding the difference between or equivalence of online and mail surveys, which demonstrates that this evidence is very inconclusive. Whereas several studies find differences—including factor structures, socially desirable answers, more extreme responses, higher item completion, higher item variability, and higher measurement error in online surveys (Klassen and Jacobs 2001; Shermis and Lombard 1999; Stanton 1998)—others indicate that online and mail surveys are equivalent (Epstein et al. 2001; Knapp and Kirk 2003). These conflicting results may be due to, for example, differences in the response processes for online and mail surveys. Because online respondents, for the most part, cannot scan, preview, review, skip, or change items, they may experience a different level of self-generated validity, which refers to the phenomenon in which responses to previous survey questions can affect answers to subsequent questions in the survey (Feldman and Lynch Jr. 1988). Moreover, computer anxiety might affect participants' responses (Buchanan and Smith 1999), or biases could occur in the way people perceive questions on screen versus on paper. Also, different screen formats and otherwise inconsistent computer administration, as well as technical or interface problems, can elicit different responses in an online survey. Finally, the print quality of paper-and-pencil surveys is still better than that of onscreen surveys, which could result in lower comprehension, less patience, and decreased reading speed on the computer (Nielsen 2000).

TABLE 3.1: Comparison of Mail and Online Surveys

	Mail	Online
Coverage	High	Low
Speed	Low	High
Control of data collection environment	Low	Low
Response rates	Low	Low
Flexibility of data collection	Low	Moderate to high
Wrong addresses	Low	High
Labor needed	High	Low
Expertise to construct	Low	Medium to high
Costs		
Invitations	Letterhead, envelopes, postage, manpower to generate, fold, stuff, and mail packets	Manpower to generate e-mails
Data entry of results	Hand data entry or scanning form Quality control checks and reentry required	Data entered by respondent. Zero data entry errors
Data handling	Error rate in hand data entry Manual processing of mailed survey response forms Results not available until final data entry and analysis is performed	Data resides on server Researcher can track results as responses are acquired
Reminders	Letterhead, envelopes, postage, manpower to generate, fold, stuff, and mail packets	Manpower to generate e-mails

Sources: Adapted from Cobanoglu, Warde, and Moreo (2001), Malhotra (1999), and Online Survey Services (<http://www.onlinesurveyservices.com/oiss/cost.htm>).

TABLE 3.2: Overview of Existing Literature that Compares Mail and Online Surveys

Author (s)	Sample	Survey topic	Survey modes	Sample size	Response (rate)	Method testing equivalence	Findings
Stanton (1998)	Professional employees	Employees' perceptions of fairness in their day-to-day interactions with their supervisors	Mail Web	N/A N/A	181 (~50%) 50 (-)	ANOVA, t-test, SEM	Differences in missing values, reported level of fairness, measurement errors. Comparable in terms of item variability.
Buchanan and Smith (1999)	Mainly students	Self-monitoring scale	Paper-and-pencil Web	N/A N/A	224 (-) 963 (-)	CFA	Comparable model fits, similar psychometric properties.
Shermis and Lombard (1999)	Organization members	Assessment of electronic media	Mail E-mail	585 585	211 (36 %) 176 (30 %)	Chi-square, multiple Mann-Whitney U test, t-test, ANOVA,	Differences in screening questions, item response rate, open-ended questions.
Cobanoglu, Warde, and Moreo (2001)	Hospitality professors	Hospitality education	Mail Fax Web	99 100 95	26 (26 %) 17 (17 %) 42 (44 %)	ANOVA	Equivalence
Epstein et al. (2001)	Students	Physical and sexual attractiveness of men and women	Paper-and-pencil Web	N/A N/A	116 (55 %) 97 (46 %)	Equivalence criterion: 20% deviation of mean score of Web from paper-and-pencil.	Equivalence in mean ratings of physical and sexual attractiveness. Difference in rating of same gender targets.
Klassen and Jacobs (2001)	Senior production or marketing professionals from a broad sample of firms	Forecasting practices	Mail Fax Disk-by-mail Web	163 161 160 110	38 (23 %) 32 (20 %) 33 (21 %) 15 (14 %)	Regression	Differences in item completion rate. Limited evidence of systematic bias in use of forecasting methods. No difference in reporting of firm-level characteristics.
Truell, Bartlett, and Alexander (2002)	Business education professionals	A dummy instrument called "Educator's survey"	Mail Web	153 153	81 (51 %) 78 (49 %)	T-test	Differences in response completeness.
Grandcolas, Rettie, and Marusenko (2003)	Students	Internet and market research	Paper-and-pencil Web	- 4000	123 (-) 157 (4%)	Chi-square, t-test, regression	For some items differences in means, skewness, and kurtosis, variance and extremity of responses.

Griffis, Goldsby, and Cooper (2003)	Middle- and senior-level managers from manufacturers and merchandisers	Logistics performance measurement	Mail Web	585 1776	59 (10 %) 254 (14 %)	t-test	Slight differences in strength with which measurement items gained support, but they were attributed to the statistical power. Consistency in conclusions.
Knapp and Kirk (2003)	Students	Mischievousness, general honesty, academic honesty, interpersonal relationship, prejudice, illegal behavior, alcohol, substance use, violence, health, sex, miscellaneous	Mail Touch-tone Web	N/A N/A N/A	174 (-) 121 (-) 57 (-)	Chi-square	Great variability in responses, but no significant difference with regard to major questions.
McDonald and Adam (2003)	Members of Australian football club	Satisfaction of members	Mail Web	1024 3200	471 (46 %) 826 (26 %)	Frequencies, Chi-square	Difference in missing values for demographic but not for content questions. Differences in variation of responses as well as mean responses. Differences in demographics.
Roster et al. (2004)	Adults in metropolitan area in southwest	Corporate/company reputation	Telephone Web	2173 974	251 (11.5 %) 272 (32.6 %)	Chi-square, t-tests	Differences in factor structures, mean agreement scores, nonresponse, and reliability.
Smither, Walker, and Yap (2004)	Managers and supervisors of a financial service company	Upward feedback ratings	Opscan Web (Intranet)	2739 2518	--- ---	Principal axis factor analysis, t-test, regression	Differences in rater and ratee characteristics. No differences in number of missing responses, factor structures, or means.

In addition to resulting in these inconclusive results, previous studies that compare online and mail surveys have methodological limitations. For example, in many cases, the results are based on small samples, target populations with strong technology familiarity (e.g., student or academic samples), or self-selected convenience samples. Knapp and Kirk (2003) use a different recruitment method; instead of sending an e-mail with the hyperlink to the Web questionnaire, they handed out an envelope containing the URL to respondents. This method requires more effort from the respondent and is more time consuming and complicated. Their survey also suffered from technical problems; the questionnaire was offline for 27 hours (Knapp and Kirk 2003). In addition, the choice of environment often acts as a limitation in previous studies on this topic. Epstein et al. (2001), for example, chose to conduct their study in a highly controlled environment. Furthermore, most studies are conducted in research areas other than marketing, such as psychology or public opinion research (e.g., Buchanan and Smith 1999; Stanton 1998). Finally, some studies provide contradictory results and are limited in both their quantity and methodological quality. Many articles compare the means of the online and offline groups, but because these means are just approximations of the underlying true population means, it is impossible to say which is better and how much importance should be attached to small but significant differences.

Despite the widespread evidence of nonequivalence, we hypothesize that data collected through online and mail surveys are equivalent. First, more recent studies have found equivalence between the two methods (Epstein et al. 2001; Knapp and Kirk 2003), which may indicate that factors such as computer anxiety or privacy concerns have been reduced as people become more familiar with the Internet. Second, over time, respondents gain increased experience with online surveys, which makes it less likely that the response process or the way people perceive questions on a screen versus on paper will cause differences. Third, though online communication in general has been found to be more open, there should be no difference in perceived anonymity between online and mail questionnaires because both are filled out in the absence of an interviewer. Fourth, the number of online panel members or e-mail addresses available in databases continues to increase steadily, ameliorating the lower coverage problems. Fifth, recommendations included in the growing body of literature on the best design of online surveys (e.g., Couper, Traugott, and Lamias 2001;

Schaefer and Dillman 1998) also has reduced problems due to different screen formats or other technical or interface problems. Therefore, we hypothesize that online and mail surveys are equivalent.

3.3 An Empirical Study

To test the equivalence between mail and online surveys, we conducted a service quality survey with a major multinational office equipment manufacturer. The data were collected in the United States, which has one of the highest Internet acceptance rates worldwide (NUA Internet Surveys 2002). Customers in the traditional paper-and-pencil group received a mailing that contained an introduction letter, the questionnaire, and a prepaid return envelope. Customers in the online group received an e-mail invitation, including a short introduction to the study with a request to participate and the hyperlink to the Web questionnaire. We avoided double entries by using a unique, eight-digit identification code for each respondent.

The items used to assess equivalence between the survey modes measured service call quality, service visit quality, and the intention to use the services of this provider again (see Table 3.3). The items were strongly driven by the SERVQUAL dimensions developed by Parasuraman, Zeithaml, and Berry (1988) and have been used in previous studies on service contact modes (Van Birgelen et al. 2002). All items were measured on a nine-point scale that ranged from "much worse than expected" (1) to "much better than expected" (9) for service call quality and service visit quality and from "very unlikely" (1) to "very likely" (9) for intentions. In addition, the questionnaire contained six open-ended questions about the nature of the respondents' complaints, additional information the respondents would like to obtain, desired improvements, positive and negative experiences, and their intention to switch to a competitor.

Participants for both the online and offline groups were recruited from the manufacturer's customer database. A stratified sampling procedure (survey mode, business units, regions, product type) was used, in which we first divided the customer database into different business units (standard and customized specialty systems), then into different regions (central, east, southwest, west), product types (low, medium, high volume machines), and finally survey modes (e-mail address is/is not available). Then, we drew a random sample from each

group to ensure that we obtained a valid and representative sample of customers.

For the mail survey, we received 694 responses, which represents a response rate of 16.58%. In the online survey, 255 customers participated (response rate 28.47%). The smaller sample for the online survey reflects the common problem that customer databases do not yet contain all customers' e-mail addresses. To rule out that potential differences in the response numbers or rates could be due to different sample characteristics, we compared the online and mail sample on several important background variables. We did not find significant differences with respect to region ($p = 0.405$), position in the company ($p = 0.133$), decision power ($p = 0.126$), preferred mode of communication ($p = 0.182$), or whether they recently had bought new products ($p = 0.224$).

3.4 Analyses and Results

To assess equivalence, we considered two specific components of data quality, namely, the completeness and accuracy of respondents' answers (Goetz, Tyler, and Lomax Cook 1984; Hansen 1980; McDaniel and Rao 1980). Completeness was assessed by comparing the number of respondents who provided answers to open-ended questions and considering the length of answers. To assess accuracy and bias differences between the online and mail surveys, we followed Hansen's (1980) suggestion to compare the distribution, or summary, of responses (means and variances) from one subgroup with the summary of another subgroup. Furthermore, we used the more rigorous, powerful, and versatile multigroup confirmatory factor analysis (CFA) approach to assess measurement invariance, which basically determines whether different survey settings produce different measures of the same attribute (Steenkamp and Baumgartner 1998).

The majority of structural equation modeling (SEM) applications in the behavioral sciences employ the maximum likelihood (ML) estimation procedure to provide parameter estimates for the hypothesized models (e.g., Bollen 1989). However, the ML estimator will exhibit desirable statistical properties (unbiased, consistent, asymptotically efficient, and approximating a χ^2 distribution) only if several important assumptions are met (Bollen 1989; West, Finch, and Curran 1995). Chief among these is the assumption that the manifest variables follow a

multivariate normal distribution. However, as in most customer satisfaction and service quality research (e.g., Brown, Churchill Jr., and Peter 1993), this assumption frequently does not hold in behavioral research.

Similarly, in this study there were significant deviations from multivariate normality in both the online and mail survey data sets, according to Mardia's (1970) and Srivastava's (1984) tests of multivariate skewness and kurtosis. When a data set deviates significantly from multivariate normality, the χ^2 statistic is inappropriate for assessing the fit of the CFA models, because it leads to an underestimation of the standard errors of the estimates (West, Finch, and Curran 1995). Several alternatives to ML estimation have been proposed for data sets that are not multivariate normal (cf. West, Finch, and Curran 1995), including the asymptotically distribution-free (ADF) estimator (Browne 1984) and the Satorra-Bentler scaled χ^2 statistic (χ^2_{SB}) with robust standard errors (Satorra and Bentler 1994). Monte Carlo simulation studies have found that the χ^2_{SB} statistic with robust standard errors outperforms the ADF estimator, especially when sample sizes are small (e.g., Chou, Bentler, and Satorra 1991). Therefore, to check for equivalence of the data sets between the online and mail surveys in this study, we employed a difference test for the χ^2_{SB} statistic, as recommended by Satorra and Bentler (2001).²

² The difference test for the Satorra-Bentler scaled χ^2 statistic can be implemented as follows:

$$\Delta\chi^2_{SB} = \frac{\Delta\chi^2}{c_d}, \text{ where } \Delta\chi^2 = \chi^2_1 - \chi^2_2, \text{ and } c_d = \frac{df_1c_1 - df_2c_2}{df_1 - df_2}, \text{ where } c_1 = \frac{\chi^2_1}{\chi^2_{SB,1}}, \text{ and } c_2 = \frac{\chi^2_2}{\chi^2_{SB,2}}.$$

TABLE 3.3: Measurement Items and Scale Reliabilities

Variable				Item				Offline		Online	
Service call quality	CR : AVE:	0.95 0.87	0.96 0.88	Competence of the [...] telephone support team.* Feedback on when your software-related problem is being resolved.* Understanding of your needs by the [...] call-handling agent. Speed of response by the [...] call-handling agent. Competence of the [...] call-handling agent.	M SD M SD M SD M SD	5.69 2.09 5.45 2.19 6.38 1.77 6.38 1.79 6.43 1.78	5.65 2.18 5.49 2.30 6.40 1.94 6.43 1.87 6.48 1.84				
Service visit	CR: AVE:	0.93 0.82	0.93 0.82	Ability of the [...] service technician to solve your problem in one visit.* The degree to which the [...] service technician provides feedback on the progress of the service visit. Competence of the [...] service technician.* Understanding of your needs by the [...] service technician. Amount of time it takes the [...] service technician to repair your equipment.	M SD M SD M SD M SD	6.59 2.07 6.71 1.94 7.04 1.86 6.97 1.85 6.56 1.92	6.57 2.09 6.68 1.98 7.04 1.86 7.01 1.85 6.57 1.96				
Intentions	CR: AVE:	0.95 0.84	0.96 0.84	I recommend [...] to someone who seeks my advice about [...]. I encourage associates, friends, and relatives to do business with [...]. I intend to do more business with [...] in the next few years. I consider [...] to be my first choice for [...].	M SD M SD M SD M SD	6.49 2.01 6.19 2.09 6.10 2.22 6.03 2.21	6.50 2.16 6.31 2.23 6.19 2.20 6.09 2.28				

Notes: CR = composite reliability, AVE = average variance extracted.

*This item was eliminated from the analysis

We used EQS 6.1 to analyze the data. We first specified a baseline CFA model for both the online and offline samples that contained 14 items. The fit indices were modest for both the offline ($\chi^2_{SB}(74) = 283.700$; normed fit index [NFI] = 0.917; confirmatory fit index [CFI] = 0.937; root mean square error of approximation [RMSEA] = 0.099) and the online ($\chi^2_{SB}(74) = 180.448$; NFI = 0.881; CFI = 0.925; RMSEA = 0.122) data. On the basis of the pattern of the standardized residuals and the modification indexes (Langrangian multiplier tests), we decided to remove four items from the analyses (two items measuring service call quality and two measuring service visit). This reduction resulted in a significantly better fit for both groups: offline $\chi^2_{SB}(32) = 77.637$; NFI = 0.984; CFI = 0.991; RMSEA = 0.045, and online $\chi^2_{SB}(32) = 84.921$; NFI = 0.959; CFI = 0.974; RMSEA = 0.081.

We evaluated the reliability of the constructs through composite scale reliability and average variance extracted (e.g., Fornell and Larcker 1981) (see Table 3.3). The composite scale reliability ranged between 0.96 and 0.98; thus, all values exceeded the cutoff point of 0.7 suggested by Nunnally and Bernstein (1994). The average variance extracted ranged from 0.89 to 0.93, so all constructs exceeded the 0.5 cutoff value proposed by Fornell and Larcker (1981). In addition, we assessed discriminant validity with a Satorra-Bentler χ^2_{SB} difference test (Satorra and Bentler 2001) and found discriminant validity for all constructs at $p < 0.001$.

Recall that to assess completeness, we compared the number of respondents who provided answers to open-ended questions and the length of the answers. For four of the six open-ended questions, there were no significant differences between the online and offline samples in terms of how many people responded ($p = 0.786, 0.864, 0.319, 0.562$). However, for the remaining two questions, we found that the online sample provided significantly more improvement suggestions (34% versus 26.7%, $p = 0.008$) and indicated whether and to which competitor the respondents wanted to switch (8% versus 4.7%, $p = 0.019$). With respect to the number of words, there again were no significant differences for four open-ended questions ($p = 0.525, 0.298, 0.233$, and 0.640). However, online respondents provided significantly lengthier improvement suggestions (mean number of words 8.16 versus 6.93, $p = 0.006$) and examples of positive experiences with the company (mean number of words 11.06 versus 8.21, $p = 0.000$).

The second aspect of response quality pertains to bias or inaccuracy (Goetz, Tyler, and Lomax Cook 1984). The results show that there are no significant differences in the means between online and offline groups ($\chi^2_{SB}(10) = 10.40$, $p = 0.406$; RMSEA < 0.001). Testing for the differences in the variance-covariance matrices ($\chi^2_{SB}(55) = 52.858$, $p = 0.557$; RMSEA < 0.001 ; CFI = 1.000; Tucker-Lewis index [TLI] = 1.001) and simultaneously for the means and the variance-covariance matrices ($\chi^2_{SB}(65) = 62.453$, $p = 0.567$; RMSEA < 0.001 ; CFI = 1.000; TLI = 1.000) also resulted in a good model fit. Because the covariances and means were invariant across the survey modes, we could pool the data from the online and offline survey, which means additional analyses to test for configural, metric, scalar, and factor covariance, factor variance, and error

variance invariance were unnecessary (Steenkamp and Baumgartner 1998; Vandenberg 2002; Vandenberg and Lance 2000).³

Therefore, we concluded that we could not identify any differences in the means and variance-covariance matrices with respect to the evaluations of service call quality, service visit quality, or intentions in the online and offline surveys. In four of the six open-ended questions, there also were no differences in the number of respondents who provided answers or in the length of their answers. Only for improvement suggestions and switching intentions did the online surveys provide slightly more responses. Online respondents also provided significantly lengthier improvement suggestions and examples of positive experiences with the company.

3.5 Discussion

ESOMAR, the World Association of Research Professionals, estimates that in 2004, 35% of market research in the United States was conducted through online surveys (ESOMAR 2004a). More and more companies now use online surveys, with their significantly lower costs, to assess service quality continuously. In the long term however, companies only profit from the cost savings of online surveys if their response quality is similar to that of mail surveys, so that outcomes from mixed-mode surveys provide a reliable basis for managers' decision making.

The selection of a data collection technique generally is based on four criteria: response rate, response bias, costs, and completion time (e.g., Malhotra 1999). In terms of the response rate, the online survey has a higher response rate than the mail survey (28.47% versus 16.58%). Even though the online sample initially was smaller, the higher response rate suggests that online surveys are preferable for companies attempting to contact busy, hard-to-reach professionals.

With respect to response characteristics, the online and mail samples produce virtually identical results. Both the composite reliability and the average

³ To rule out that the underlying factor structures might be different, we assessed measurement invariance following the procedure recommended by Vandenberg and Lance (2000). To examine invariance—starting with the analysis of configural, then metric, scalar, and factor covariance and finally factor variance invariance—we tested increasingly restrictive hypotheses. In line with Steenkamp and Baumgartner (1998), we also tested for error variance invariance. Our results indicate that the data fit well and that the Satorra-Bentler Chi-square difference test (Satorra 2001) was not significant ($p \geq 0.124$). For more information on testing measurement invariance, see Steenkamp and Baumgartner (1998) and Vandenberg and Lance (2000).

variance extracted show consistently high levels for both groups. The number and length of responses to open-ended questions also demonstrates similar results for the online and mail surveys. In terms of accuracy, we find that the means and the variance-covariance matrices are equal across modes. Thus, the results from online and offline surveys are comparable and produce equally usable data. The only minor differences are that online respondents provide more improvement suggestions, indicate more often to which competitor they want to switch, and provide lengthier answers about their positive experiences with the company and improvement suggestions. A possible explanation for this outspokenness by online respondents may be the reduced social context information on the Internet. Persons often are sensitive to the variables that can influence the content of a conversation and inhibit or facilitate what is said, how, and by whom. Empirical evidence from e-mail communications (Sproull and Kiesler 1986) suggests that social context information in online surveys is weak, which increases the respondents' perceived anonymity and produces relatively self-centered and unregulated behavior. Respondent behavior thus is likely to be more extreme, more impulsive, and less socially differentiated, because these people are relatively unconcerned about making a good appearance (Sproull and Kiesler 1986).

Unfortunately, we did not record information about data collection costs and response times, the third and fourth criteria. However, several studies and meta-analyses suggest that online surveys are faster and cheaper (e.g., Illieva, Baron, and Healey 2002). The findings of equivalence between online and mail surveys support the use of online surveys, whose response quality is comparable to that of mail surveys while their response time is faster and their costs are lower. Equivalence between online and mail surveys is extremely compelling for companies that rely on continuous measurements of service quality. For example, decision-support tools use continuous input from customers to predict future behavior and revenues. Because customer input enables predictions about crucial factors such as future revenues, it is absolutely necessary that surveys produce timely and reliable results at a low cost, as can be achieved through online surveys.

However, even though the study supports the equivalence of online and mail surveys, we recognize that the adequacy of online surveys depends mainly on the strength of the online sample. In this study, the online and mail samples

were comparable; thus, we could focus on examining whether there were any systematic differences in response quality or measurement invariance. We did not find any such differences, which is a promising result. However, differences between online and mail surveys might occur because of incomparable samples. For online surveys, we recommend that users carefully examine the sample frame before the data are collected. Because no post hoc comparisons are possible, the reliability and validity of a pure online survey stand and fall with the representativeness of the sample. In a B2B context, many companies maintain a database from which they can draw a random sample, but most B2C companies must rely on e-mail addresses obtained from online panels. Although most online panels are very large and employ sophisticated weighting techniques, they may not be able to replicate results from more traditional research methods (Sparrow and Curtice 2004). Furthermore, empirical evidence suggests that no simple weighting factor or adjustment strategy can make on- and offline samples comparable (e.g., Vehovar, Lozar Manfreda, and Batagelj 1999).

If only a limited number of e-mail addresses are available, mixed-mode surveys might be a viable alternative to contact those members for whom an e-mail address is available online while surveying others with a regular mail survey (Dillman 2000). Using mixed-mode surveys can increase both response rates and the representativeness of the sample and maintain the online survey benefits of faster response times and lower costs. We demonstrate how the strength of the online sample can be assessed through a comparison, with the mail survey, of relevant sample characteristics. Thus, the response quality of and equivalence between online and mail surveys can be tested easily by comparing (1) their completeness and accuracy according to the number and length of open-ended questions, (2) the distribution of responses by considering the means and variances, and (3) measurement invariance.

3.6 Limitations and Future Research Guidelines

The study was conducted within only one country, the United States. Additional studies should examine whether the findings hold true for other countries, especially those with lower rates of Internet adoption. Moreover, we compared online and mail surveys because the type of research for which mail surveys are used is most likely to be supplemented by online surveys. However, it might be interesting to examine how alternative modes, such as telephone

surveys, compare to online surveys. Further studies could also examine the impact of any potential moderating factors on the results.

In addition, the quality of online samples must receive further attention, though we did not find systematic biases between online and mail surveys. Also, the customers in the study are not end consumers but rather business customers; the findings for end customers might differ from those presented herein. According to the company, the small number of e-mail addresses in the database is due to erratic updates by the sales force, not a systematic bias in, for example, the innovativeness of those customers for whom an e-mail address is available. However, a selection bias cannot be ruled out completely because the availability of e-mail addresses might contain an element of self-selection.

The quality of online studies depends on the quality of the online sample. In general, the best way to secure representative results is by maximizing response rates (Hansen 1980) to reduce nonresponse error. However, this conventional wisdom may not be true for online panels, in which many respondents are members of several panels or participate because of the money they can win (e.g. "The free Get Paid to Take Surveys Online Guide;" <http://www.surveys4money.com/>). A large response rate also could indicate that more highly motivated, survey-prone, or professional respondents have participated. Even if the demographic profile is representative, it is reasonable to assume that the responses are not in such a case. Insights into these issues will advance knowledge about the quality of online surveys and thereby help empirically determine the potential of Internet-based research.

CHAPTER 4: COMPARING THE GENERALIZABILITY OF ONLINE AND MAIL SURVEYS IN CROSS-NATIONAL SERVICE QUALITY RESEARCH⁴

To compare the generalizability of online and mail surveys in a cross-national service quality study, the authors use G-theory and find a comparable level of generalizability, though online surveys benefited from considerably lower costs. This study contributes to the current comparison of the response quality between online and mail surveys. Furthermore, the authors illustrate how G-theory can be used to compare online and mail surveys and take data collection costs into account. Important implications include the process and results of comparing two survey modes and the effects for service research.

⁴ This chapter is largely based on Deutskens, Elisabeth C., Ad de Jong, Ko de Ruyter, and Martin G.M. Wetzels (2006), "Comparing the Generalizability of Online and Mail Surveys in Cross-National Service Quality Research," *Marketing Letters*, Forthcoming.

4.1 Introduction

Across a wide variety of domestic and international markets, an increasing number of firms have recognized that service excellence drives competitive advantage. Therefore, customer perceptions of service quality have become a strategic priority, and many companies now collect these customer perceptions on an ongoing basis. However, the continuous measurement of service quality, particularly cross-nationally, is costly and time consuming, which has led more and more firms to resort to online surveys.

According to ESOMAR, the World Association of Research Professionals, online surveys accounted for 35% of the U.S. survey research market at the end of 2004 (ESOMAR 2004a). Online surveys allow easier access to a wider range of customers, irrespective of their geographical location (e.g., Illieva, Baron, and Healey 2002; Johnson 2001). Furthermore, they may reduce field costs by 50–80%, lower response and data processing times substantially, and enable the centralization of research project management (e.g., Dillman 2000; Illieva, Baron, and Healey 2002). Although extensive evidence details response rates and completion times in online surveys (e.g., Deutskens et al. 2004; Illieva, Baron, and Healey 2002), evidence about their quality seems sparse and inconclusive (e.g., Epstein et al. 2001; Stanton 1998). Several authors have found, for example, that online and mail surveys produce different results (e.g., McDonald and Adam 2003; Shermis and Lombard 1999) that cannot be equated through a simple weighting factor or adjustment strategy (e.g., Robinson, Neustadt, and Kestnbaum 2002; Vehovar, Lozar Manfreda, and Batagelj 1999).

In particular, virtually no survey research addresses the required or desired response quality in relation to data collection costs, despite the importance of the response quality/cost ratio in trade-off decisions. Response quality, which is generally defined as the effectiveness, accuracy or relevancy of answers, and costs are directly linked (e.g., if more items are added to a questionnaire to improve the reliability of the constructs and thus the response quality, the cost must also increase). Surveyors can thus pursue two different goals: (1) maximize response quality given a certain cost level or (2) minimize costs given the desired response quality level. Because research resources are invariably

limited and the costs of large-scale international performance measurement, even when administered online, may still be substantial, it seems relevant to evaluate how these restricted research resources may best be employed. Therefore, the aim of this study is to address response quality and data collection costs of online and mail surveys simultaneously in the context of a cross-national service quality survey using generalizability theory (G-theory).

G-theory presents an elaborate psychometric method for an appealing and complete approach. In the first step of a G-study, the researcher generalizes from observations of a given measurement condition to a set of different conditions to determine the reliability of the measure. G-theory can also be used to determine the most cost-effective research design for a required or desired level of generalizability, including the number of scale items and respondents necessary (e.g., Finn and Kayandé 1997). This information can be used in the design of a subsequent decision study (D-study) so that the desired or required level of generalizability can be achieved at minimal cost.

This study contributes to the survey literature in three important ways. First, we add to the current discussion about the comparison of response quality between online and mail surveys by providing evidence about the generalizability of both types. Second, we illustrate how G-theory can compare online and mail surveys while taking data collection costs into account, which is especially important in light of the recent trend toward mixed-mode surveys as a means to improve response rates while minimizing response times and data collection costs (e.g., Dillman 2000; Roster et al. 2004). Third, we provide managerial guidelines pertaining to the combined use of online and mail surveys to optimize survey costs while maintaining a high response quality.

4.2 Literature Review

4.2.1 Online and Mail Surveys

The exponential growth of the Internet has induced one of the most profound developments in survey methodology (e.g., Dillman 2000), namely, the collection of survey data via the Internet. Online surveys are increasingly used in both academic (e.g., Iqbal, Verma, and Baran 2003; Lynch and Ariely 2000; Mandel and Johnson 2002; Meuter et al. 2000; Novak, Hoffman, and Yung 2000; Toubia, Hauser, and Simester 2004) and market (e.g., Hogg 2002) research. Persuasive arguments for choosing online surveys over traditional methods

include benefits such as lower costs, faster response times, and wider geographic reach, which makes them especially suitable for cross-national research (e.g., Dillman 2000; Illieva, Baron, and Healey 2002; Johnson 2001; Roster et al. 2004). Furthermore, proponents of online surveys argue that the Internet allows for the use of uncomplicated directions (e.g., through automatic routing), as well as richer and more interesting question formats (Klassen and Jacobs 2001; Simsek and Veiga 2001). Online surveys also have been found to be useful in reaching busy professionals, a population for whom mail surveys suffer from low and continually declining response rates. Yet, in terms of coverage, addressing accuracy, and representativeness, mail surveys are still preferred. In Table 4.1, we provide an overview of the strengths and weaknesses of both methods, as have been identified in the literature.

TABLE 4.1: Comparison of Mail and Online Surveys

	Mail	Online
Coverage	High	Low
Speed	Low	High
Control of data collection environment	Low	Low
Response rates	Low	Low
Flexibility of data collection	Low	Moderate to high
Wrong addresses	Low	High
Labor needed	High	Low
Expertise to construct	Low	Medium to high

Sources: (Cobanoglu, Warde, and Moreo 2001; Malhotra 1999; OnlineSurveyServices).

As can be seen in Table 4.1, both methods have certain advantages but disadvantages as well. To help researchers evaluate the relative strengths and weaknesses of each method and the selection of a data collection technique, Wiseman (1972) has proposed four criteria: response rate, completion time, response quality, and costs. Response rate and completion time have already been examined; the evidence indicates that response times in online surveys are considerably shorter than for mail surveys, whereas response rates seem higher in mail surveys (e.g., Cobanoglu, Warde, and Moreo 2001; Deutskens et al. 2004; Illieva, Baron, and Healey 2002; Roster et al. 2004).

Response quality and data collection costs have received the least research attention, but they constitute important measures for two main reasons. First, response quality and costs might be considered the most relevant issues for researchers. The ultimate goal of every study is to collect high-quality data that provides a reliable basis for decision making. At the same time, the advance of

online surveys and increasing worldwide competition have amplified the opportunities for small companies to conduct surveys and thus put marketing researchers under enormous cost pressures. Second, as we mentioned previously, response quality and data collection costs are directly linked, but they have not been examined for online versus mail surveys.

Some studies have compared online and mail surveys with the aim of assessing the superiority of one method over the other. Whereas some studies find no significant differences between online and mail surveys (Epstein et al. 2001; Knapp and Kirk 2003), others ascertain an increased amount of socially desirable answers and more extreme responses, greater item completion, higher item variability, and higher measurement error in online surveys (Klassen and Jacobs 2001; Shermis and Lombard 1999; Stanton 1998). Dillman (2000) posits that possible reasons for these differences could be that respondents to online surveys have a different process of responding and are thus not able to scan, preview, review, skip, or change items. Moreover, computer anxiety might affect participants' responses (Buchanan and Smith 1999), or biases could occur in the way people perceive questions onscreen versus on paper (Dillman 2000). Couper (2000) also suggests that different screen formats or otherwise inconsistent computer administration, as well as technical and interface problems, can elicit different responses to an online survey.

Practitioners have resorted to post-stratification or propensity score weighting to deal with generalizability issues of online surveys, even though empirical evidence suggests that no simple weighting factor or adjustment strategy can make on- and offline samples comparable (e.g., Robinson, Neustadt, and Kestnbaum 2002; Vehovar, Lozar Manfreda, and Batagelj 1999). Thus, empirical evidence from studies comparing online and mail surveys clearly is mixed. Furthermore, and perhaps more important, most studies have compared online and mail surveys in terms of data superiority. But what does "superiority" imply if we find, say, more extreme responses by online respondents? Because the means from both online and mail surveys are just approximations of the underlying true population mean, it is impossible to say which is better and how much importance should be attached to small but significant differences. What researchers are ultimately interested in is obtaining high response quality, which is defined as effective, accurate, and relevant

answers that can be generalized to other settings. However, the generalizability of online and mail surveys has not been examined empirically.

In addition, data collection costs have rarely been discussed in relation to the response quality of online and mail surveys. Claims regarding the cost effectiveness of online surveys in international service research are based largely on anecdotal evidence and have not been systematically analyzed relative to mail surveys, which is surprising because response quality and data collection costs are interdependent. That is, minimizing the number of items in a questionnaire reduces its overall length and thus respondent fatigue and the costs of the study (Sharma and Weathers 2003), but it interferes with the reliability of the results. Researchers are forced to either minimize costs at the expense of generalizability or maximize generalizability at the expense of greater costs. We therefore attempt to evaluate how researchers can best use their limited resources to attain acceptable response quality.

4.3 Research Setting

We evaluate the performance of mail and online surveys in a customer-perceived service quality setting using an adapted SERVQUAL scale. Service quality presents an excellent context for several reasons. First, it is one of the most prominent research areas in marketing, which makes this study interesting for a wide audience. Second, because of the continuous tracking of service quality, companies could benefit enormously from the significantly lower costs of online surveys. If the generalizability of online and mail surveys is comparable and no mode differences exist, companies would profit. Third, the SERVQUAL dimensions have been used in previous G-studies (Finn and Kayandé 1997), though not to compare online and offline surveys across countries.

We consider G-theory an appropriate framework in this context because it has been identified as a comprehensive psychometric method for assessing the consistency of performance instruments (Rentz 1987). The basic concept underlying G-theory is that researchers generalize from the observation at hand to some class of observations as a mean to determine the reliability of a measure (Brennan 2001; Cronbach et al. 1972; Cronbach, Rajaratnam, and Gleser 1963; Peter 1979; Rentz 1987, 1988). For example, one might determine the reliability of a service quality measure by generalizing from observations collected on a particular occasion (e.g., 50 customers surveyed via mail in January last year) to

a set of different measurement conditions (e.g., different items, survey modes, countries, points in time) (Rentz 1987). In this context, for example, companies might want to compare how customers in different countries perceive their service quality or how the importance of each service quality aspect varies across countries (Finn and Kayandé 1997).

In contrast to classical test theory, G-theory explicitly recognizes multiple sources of variance, or *facets*. We distinguish between two different types of facets, namely, of generalization and differentiation. In generalizability terminology, a *facet of generalization* refers to a set of conditions that contribute unwanted variation (measurement error) to the observations in a study (Rentz 1987, p. 21). Thus, a facet of generalization represents a factor across which the researcher wants to generalize (e.g., respondents, items) (Finn and Kayandé 1997). In contrast, a *facet of differentiation* is a set of objects that the researcher wants to compare using the measurement instrument (Rentz 1987). Therefore, the facet of differentiation is also called the object of measurement. In most studies, the object of measurement pertains to a single facet (e.g., countries, business units), though there can also be interaction between two facets (e.g., countries \times business units).

The design of a G-study is comparable to the design of an experiment. In the first step, the researcher receives measurements for the facets of differentiation in the various conditions for all the relevant facets of generalization. Facets may be crossed or nested, random or fixed. Next, a generalizability coefficient can be computed for each facet or combination of facets of generalization. The coefficient indicates the extent to which the observations can be generalized to the set of measurement conditions of interest (Rentz 1987, p. 20f; 1988, p. 142). The strength of G-theory is that it takes into account the different purposes of measurement. In this context, this strength enables us to compare the assessments of service performance through online and mail surveys for different research questions. In the context of a cross-national customer satisfaction survey conducted through online and mail surveys, we could identify five factors per survey mode that contribute to the variability of service quality, namely: (1) countries, (2) business units, (3) persons, (4) aspects, and (5) items. Finally, a D-study collects data for a particular purpose (Rentz 1987). In this case, the D-study is designed using the

information from the G-study, thereby optimizing data collection costs while employing a design that produces the desired level of generalizability.

As we mentioned previously, researchers can either maximize the response quality given a certain cost level or minimize costs given a desired level of generalizability. In either case, researchers can identify the most efficient design: The generalizability criterion is satisfied at the lowest cost. Although the benefits of the G-theory approach for optimizing designs for marketing measures have been explicitly acknowledged, marketing scholars and practitioners have rarely adopted it to minimize the costs of surveys. This feature becomes especially interesting, however, for mail and online surveys because, in contrast to mail surveys, online surveys have a very appealing cost structure with extremely low variable costs. Even though the fixed costs may be slightly higher, online surveys are less expensive than mail surveys overall. To illustrate this claim, we conducted in-depth interviews with five researchers working at three different international market research agencies to develop two cost functions, C1 and C2, for mail and online surveys, respectively. More details on how the cost functions were derived can be found in Table 4.2.

TABLE 4.2: Decomposition of Cost Functions for Mail and Online Surveys

Costs	Mail	€	Online	€
Fixed costs	intellectual effort	770.00	intellectual effort	770.00
	sampling	220.00	hosting	500.00
	coding open questions	330.00	batch creating	110.00
	scripting (layout)	500.00	sampling	220.00
			coding open questions	330.00
			scripting effort	700.00
		1820.00		2630.00
Variable costs per send out questionnaire	printing	0.20		----
	envelopes	0.30		
	handling	0.45		
	stamps	0.40		
		1.35		
Variable costs per returned questionnaire	return costs	0.79	hosting	0.02
	data entry	0.65		
		1.44		
Lengthening by one additional question	additional printing costs	0.01		5.00
	more expensive stamps	0.01		
	additional costs per additional item per send out questionnaire	0.02		
	additional data entry	0.02		
	more expensive return stamps	0.02		
	additional costs per additional item per returned questionnaire	0.04		

*The numbers are based on interviews with market research experts from large market research agencies but are, of course, only a suggestion and not universally valid. Optimization can easily be repeated with different inputs for the cost function.

$$(C1) \quad \begin{aligned} C_T &= C_F + C_S + C_R + C_Q, \text{ with } C_S = N * V_S, C_R = n * V_R, C_Q = Q * N * V_{QS} + Q * n * V_{QR}, \text{ and } n = N * R \\ \text{or} \\ C_T &= C_F + (N * V_S) + (N * R * V_R) + (Q * N * V_{QS} + Q * N * R * V_{QR}) \\ &= C_F + N(V_S + R * V_R + Q * V_{QS} + Q * R * V_{QR}), \end{aligned}$$

where C_T = total estimated survey costs; C_F = fixed costs; C_S = total direct costs of mailing N questionnaires; C_R = total direct costs of receiving returned questionnaires; C_Q = total direct costs of additional questions; V_S = estimated variable costs of mailing one questionnaire; V_R = estimated variable costs of receiving one returned questionnaire; V_{QS} = estimated variable costs per invitation of extending the questionnaire by one additional question; V_{QR} = estimated variable costs per respondent of extending the questionnaire by one additional question; Q = number of questions; R = response rate; and n = required number of replies. Furthermore,

$$(C2) \quad \begin{aligned} C_T &= C_F + C_Q + C_H, \text{ with } C_Q = Q * V_Q, \text{ and } C_H = n * V_H, \text{ and } n = N * R \\ \text{or} \\ C_T &= C_F + Q * V_Q + N * R * V_H, \end{aligned}$$

where C_Q = total direct costs of programming; C_H = total direct costs of hosting; V_Q = estimated variable costs of programming one question; V_H = estimated variable costs of hosting one additional respondent; Q = number of questions; R = response rate, and N = number of e-mail invitations sent.

Using the experience and internal records of the marketing research agencies that assisted us with input for these cost functions, we arrive at the following cost functions for mail and online surveys, respectively:

$$C_{T \text{ Mail}} = \text{€}1820 + \text{€}1.35 * N + \text{€}1.44 * N * R + \text{€}0.015 * Q * N + \text{€}0.04 * Q * N * R$$

$$C_{T \text{ Online}} = \text{€}2630 + \text{€}5 * Q + \text{€}0.02 * N * R.$$

To identify the most efficient research design and in line with previous generalizability studies (e.g., Finn and Kayandé 1997; Sanders, Theunissen, and

Baas 1989), we use the branch-and-bound algorithm (Salkin 1975) to solve the optimization problem by minimizing the cost of the research design given a desired level of the G-coefficient $E\hat{p}^2$. In the next section, we present the empirical study, in which we compare the generalizability of a cross-national investigation of service quality for online and mail surveys. We are specifically interested in identifying the least expensive measurement design that allows a reliable assessment of the object of measurement (service quality).

4.4 An Empirical Demonstration

4.4.1 Research Design and Data Collection

The data for this study were obtained in 2002 in the United States and United Kingdom through a service quality survey of (business) customers of a major multinational office equipment manufacturer. The quality aspects we investigated were telephone support, registering a service call, service visits, complaint handling, products, and ordering supplies. The multi-item scale used to measure the level of service quality of these aspects is based on the SERVQUAL dimensions developed by Parasuraman, Zeithaml, and Berry (1988). Because the research context is a continuous evaluation of service quality, the questionnaire has been used in its current form for several years and only adapted to changes in the service offering. All items were measured with a nine-point scale ranging from "much worse than expected" (1) to "much better than expected" (9). The scales revealed Cronbach's alphas between .80 and .96, providing clear evidence of construct reliability (Nunnally and Bernstein 1994). Table 4.3 provides an overview of the measurement instrument and respective reliabilities.

Participants for both the online and offline surveys were recruited from the manufacturer's customer database. A stratified sampling procedure (business units, regions, product type) was used for both samples, followed by a random sample drawn from each group to ensure a valid and representative sample of customers was obtained. Customers in the traditional paper-and-pencil group received a mailing containing an introduction letter, the questionnaire, and a prepaid return envelope. Customers in the online group received an e-mail invitation, including a short introduction with a request to participate and the hyperlink to the Web questionnaire. We avoided double entries by using a unique eight-digit ID for each respondent.

TABLE 4.3: Measurement Items, Descriptives, and Scale Reliabilities

Variable	Item	Online			Offline		
		N	M (SD)	Reliability	N	M (SD)	Reliability
Telephone support	Ability of the [...] telephone support team in solving your software-related problem in one phone call.	USA	5.38 (2.38)	0.971	402	5.24 (2.19)	0.959
		UK	5.20 (1.94)	0.952	149	5.21 (1.90)	0.953
	Competence of the [...] telephone support team.	USA	5.65 (2.18)		401	5.69 (2.10)	
		UK	5.40 (1.99)		148	5.50 (1.89)	
	Understanding your needs by the [...] telephone support team.	USA	5.61 (2.22)		401	5.69 (2.09)	
Registering a service call	Speed of response by the [...] call-handling agent.	UK	5.38 (1.83)		147	5.48 (1.93)	
		USA	6.43 (1.87)	0.925	836	6.38 (1.79)	0.929
	Competence of the [...] call-handling agent.	UK	5.86 (1.34)	0.818	354	5.76 (1.38)	0.847
		USA	6.48 (1.84)		833	6.43 (1.78)	
	Service visit	The degree to which the [...] service technician provides feedback on the progress of the service visit.	UK	5.91 (1.52)		355	5.95 (1.58)
USA			6.68 (1.98)	0.944	900	6.71 (1.94)	0.924
Availability of spare parts during the service visit.		UK	6.13 (1.67)	0.813	331	6.19 (1.63)	0.902
		USA	6.30 (2.07)		890	6.21 (2.05)	
Competence of the [...] service technician.		UK	5.01 (1.85)		326	5.56 (1.77)	
Complaint handling	Manner in which the [...] service technicians present themselves.	USA	7.04 (1.86)		901	7.04 (1.86)	
		UK	6.39 (1.55)		329	6.42 (1.63)	
	Accessibility of [...] when communicating your complaint or inquiry.	USA	7.24 (1.71)		903	7.21 (1.74)	
		UK	6.37 (1.52)		332	6.32 (1.46)	
	Time taken by the [...] staff in handling your complaint or inquiry.	USA	5.16 (1.98)	0.947	240	5.25 (1.86)	0.947
Products	Ability of the [...] staff in handling your complaint or inquiry correctly the first time.	UK	4.28 (2.04)	0.940	91	4.12 (2.03)	0.955
		USA	4.95 (2.20)		240	5.12 (2.13)	
	Competence of the [...] staff in handling your complaint or inquiry.	UK	3.70 (2.08)		90	4.02 (2.15)	
		USA	4.42 (2.42)		235	4.48 (2.46)	
	Features of the product.	UK	3.54 (2.26)		91	3.62 (2.20)	
Copy / print quality of the product.	Speed of response by the [...] call-handling agent.	USA	4.81 (2.38)		237	4.95 (2.24)	
		UK	4.24 (2.34)		90	3.92 (2.23)	
	Competence of the [...] staff in handling your complaint or inquiry.	USA	6.34 (1.74)	0.882	916	6.22 (1.54)	0.852
		UK	5.86 (1.29)	0.802	386	5.67 (1.32)	0.819
	Copy / print quality of the product.	USA	6.64 (1.70)		918	6.50 (1.66)	
UK		6.19 (1.42)		394	5.84 (1.47)		

Ordering supplies	Responsiveness of [...] when ordering supplies.						
	USA	278	6.61 (1.54)	0.932	988	6.55 (1.60)	0.925
Receipt of the supplies in an undamaged condition.	UK	107	5.67 (1.61)	0.840	352	5.56 (1.50)	0.861
	USA	276	6.74 (1.56)		988	6.84 (1.60)	
Quality of the supplies.	UK	106	6.20 (1.39)		355	5.99 (1.50)	
	USA	279	6.81 (1.53)		989	6.83 (1.57)	
	UK	107	6.17 (1.36)		358	6.04 (1.40)	

In total, 10,506 questionnaires were sent to customers, of which 1957 usable questionnaires were returned (18.63%). Responses per country were as follows: U.S. offline, 1000 of 6030 (16.58%); online, 377 of 1324 (28.47%); U.K. offline, 439 of 2622 (16.74%); and online, 141 of 530 (26.60%). The smaller sample for the online survey reflects the common problem that the customer database does not yet contain all customer e-mail addresses, which, according to the company, is due to erratic updates by the sales force, not to systematic bias (e.g., the innovativeness of those customers for whom an e-mail address is available).

Also, customers in this case are not end users but business customers. To rule out the possibility that response differences could be due to different sample characteristics, we compared the online and mail samples on several important background variables. We did not find significant differences with respect to position in the company ($p = .111$), business in which the company operates ($p = .666$), role in the decision process for new products ($p = .328$), how new products and services are obtained ($p = .969$), or whether the products are connected to a computer network ($p = .596$).

Both off- and online data sets have a partially hierarchical structure; subjects are nested within countries (U.S. and U.K.) and business units (standard systems and customized specialty systems). Furthermore, the data are unbalanced because (1) the number of subjects is unequal for the different countries and business units and (2) the number of items is unequal for the different aspects. Also, not every respondent had utilized all service aspects, and screening questions before every section ensured that respondents would not answer questions about aspects he or she had not experienced. Unbalanced data comprise statistical and conceptual complexities, which is especially problematic for D-studies because general procedures for unbalanced data are not yet available (Brennan 2001). In contrast, balanced designs benefit from clear-cut research designs and simplified estimations of variance components and other parameters; they also enable a straightforward comparison of G- and D-study results. Given the cross-national focus, we decided to balance the number of subjects per country by using the country with the smallest number of respondents per business units as a base. Therefore, we randomly selected 126 customers per country for the G-study (see Kane, Gillmore, and Crooks 1976), which resulted in a more balanced data set with an equal number of respondents

per country. To estimate the variance components, we used the minimum variance quadratic unbiased (MIVQUE) estimation method (Hartley, Rao, and LaMotte 1978), which is especially suitable for balanced designs.

4.4.2 Generalizability Analysis

Table 4.4 provides detailed information about the experimental structure of the G-study designs. In design 1, country was specified as the facet of differentiation, whereas in design 2, both country and aspect were specified as facets of differentiation. The designs that use business units and business units and aspects as the facets of differentiation have similar structures.²

The structure of the data enables us to estimate a restricted variance model, which consists of 14 variance components (see Table 4.5), which we use subsequently to calculate the G-coefficient for the different facets of differentiation. The estimated variance components for the mail and online survey are reported in column 1 and 2 respectively. The associated percentages of variance for each mode are shown in column 3 and 4.

The results show that in the mail survey, the main sources of variance are persons (37.02%), random error (27.62%), and the interaction term persons × aspects (26.41%). Smaller but still considerable variance components derive from aspects (5.19%) and the countries × aspects interaction (2.12%). Business units × aspects only contributes .33% to the overall variance.

Similarly, the major sources of variance in the online survey appear to be random error (35.51%), persons (30.39%), and persons × aspects (17.73%). Furthermore, sizeable percentages of variances are attributed to aspects (5.10%), countries × business unit (3.13%), countries × aspects (2.47%), and countries (3.56%). Business units × aspects only contributes .86% to the overall variance. Thus, the results demonstrate that online and mail surveys have comparable variance components. Only the interaction of country × aspect, the facet of differentiation in the second design, constitutes a substantially larger percentage of variance in the online survey.

² We also estimated a third design in which the interaction term country × aspect (business unit × aspect) remains the object of measurement, but business units (countries) are specified as a fixed facet. The country × business unit × aspect interaction has zero variance, and therefore, the results of design 2 and 3 are identical.

TABLE 4.4: Different Measurement Designs for Countries (Similar for Business Units)

DESIGNS	DESIGN 1	DESIGN 2
Facets of differentiation	Countries (C), random	Countries (C), random Aspects (A), random
Facets of generalization	Business units (B), random Persons (P), random Aspects (A), random Items (I), random	Business units (B), random Persons (P), random Aspects (A), random Items (I), random
Nesting specifications	P nested within C and B I nested within A	P nested within C and B I nested within A
Sources of variance	C, B, P:CB, A, I:A, CB, CA, CI:A, BA, BI:A, PA:CB, PI:CBA, CBA, CBI:A	C, B, P:CB, A, I:A, CB, CA, CI:A, BA, BI:A, PA:CB, PI:CBA, CBA, CBI:A
Universe score variance components ($\hat{\sigma}_i^2$)	$\hat{\sigma}_c^2$	$\hat{\sigma}_{ca}^2$
Relative error variance components ($\hat{\sigma}_e^2$)	$\begin{aligned} &= \frac{\hat{\sigma}^2(p:cb)}{n_{p:b}} + \frac{\hat{\sigma}^2(cb)}{n_b} + \frac{\hat{\sigma}^2(ca)}{n_a} \\ &+ \frac{\hat{\sigma}^2(ci:a)}{n_{ai}} + \frac{\hat{\sigma}^2(pa:cb)}{n_{bpa}} \\ &+ \frac{\hat{\sigma}^2(pi:cba)}{n_{bpai}} + \frac{\hat{\sigma}^2(cba)}{n_{ba}} \\ &+ \frac{\hat{\sigma}^2(cbi:a)}{n_{bia}} + \frac{\hat{\sigma}^2(rand.error)}{n_{bpaal}} \end{aligned}$	$\begin{aligned} &= \frac{\hat{\sigma}^2(ci:a)}{n_i} + \frac{\hat{\sigma}^2(pa:cb)}{n_{bp}} \\ &+ \frac{\hat{\sigma}^2(pi:cba)}{n_{bpi}} + \frac{\hat{\sigma}^2(cba)}{n_b} \\ &+ \frac{\hat{\sigma}^2(cbi:a)}{n_{bi}} + \frac{\hat{\sigma}^2(rand.error)}{n_{bpi}} \end{aligned}$

TABLE 4.5: Variance Components and G-Coefficients Based on Current Designs

Source of Variation	G-Study		Current D-study design					
			2		6		3	
$\eta_{\text{countries}} =$	1		2		6		3	
$\eta_{\text{business units}} =$	1		2		6		3	
$\eta_{\text{persons}} =$	1		63		6		3	
$\eta_{\text{aspects}} =$	1		6		6		3	
$\eta_{\text{items}} =$	1		3		3		3	
Source of Variation	Estimated Variance Components		Estimated Variance Components		Estimated Variance Components		Estimated Variance Components	
	Mail Survey	Online Survey	Mail Survey	Online Survey	Mail Survey	Online Survey	Mail Survey	Online Survey
Countries	0.02802	0.17817	0.60%	3.56%	0.01401	0.08909	0.01401	0.08909
Business units	0.00496	0.00000	0.11%	0.00%	0.00248	0.00000	0.00248	0.00000
Persons	1.72712	1.52264	37.02%	30.39%	0.00685	0.00604	0.00685	0.00604
Aspects	0.24220	0.25568	5.19%	5.10%	0.04037	0.04261	0.04037	0.04261
Items	0.00988	0.03978	0.21%	0.79%	0.00055	0.00221	0.00055	0.00221
Countries x Business units	0.00000	0.15684	0.00%	3.13%	0.00000	0.03921	0.00000	0.03921
Countries x Aspects	0.09872	0.12375	2.12%	2.47%	0.00823	0.01031	0.00823	0.01031
Countries x Items	0.00000	0.00000	0.00%	0.00%	0.00000	0.00000	0.00000	0.00000
Countries x Business units x Aspects	0.00000	0.00000	0.00%	0.00%	0.00000	0.00000	0.00000	0.00000
Countries x Business units x Items	0.01235	0.00000	0.26%	0.00%	0.00017	0.00000	0.00017	0.00000
Business units x Aspects	0.01548	0.04327	0.33%	0.86%	0.00129	0.00361	0.00129	0.00361
Business units x Items	0.00586	0.02283	0.13%	0.46%	0.00016	0.00063	0.00016	0.00063
Persons x Aspects	1.23198	0.88856	26.41%	17.73%	0.00081	0.00059	0.00081	0.00059
Persons x Items	0.00000	0.00000	0.00%	0.00%	0.00000	0.00000	0.00000	0.00000
Random error	1.28868	1.77922	27.62%	35.51%	0.00028	0.00039	0.00028	0.00039
Total variance	4.66525	5.01074	100.00%	100.00%	0.07521	0.19469	0.07521	0.19469

Part B

	Estimated Variance Components.		Expected Variance Components	
	Mail Survey	Online Survey	Mail Survey	Online Survey
Main Effects:				
Rel. error var. for countries	4.35885	4.47101	0.01635	0.05654
G-coefficient for countries	0.00639	0.03832	0.46145	0.61172
Rel. error var. for business units				
G-coefficient for business units	2.66295	2.89741	0.01150	0.01774
Rel. error var. for aspects	0.08337	0.08109	0.77830	0.70604
G-coefficient for aspects				
If countries fixed and business units are fixed:				
Rel. error var. for aspects	2.54875	2.73039	0.00198	0.00382
G-coefficient for aspects	0.12268	0.13406	0.96178	0.93664
Interaction Effects:				
Rel. error var. for countries x aspects	2.53301	2.66778	0.00127	0.00098
G-coefficient for countries x aspects	0.03751	0.04433	0.86623	0.91322
Rel. error var. for business units x aspects	2.53887	2.69061	0.00143	0.00161
G-coefficient for business units x aspects	0.00606	0.01583	0.47371	0.69078

Table 4.5 also shows the results of the D-study, which is based on the research design characteristics (i.e., two countries, two business units, 63 customers per country per business unit, six aspects, and three items per aspect). In line with the G-study, we assume a balanced D-study. If we assume that all facets are random, the G-coefficient for the mail survey is .87 for the comparison of countries across different aspects; for the online data, the G-coefficient is .91 (Part B Table 4.5). In addition, comparing business units across different aspects yields a G-coefficient of .47 for the mail survey and .69 for the online survey. Small G-coefficients indicate that additional sampling is required, which makes it more costly to produce the required level of generalizability.

Furthermore, to estimate the accuracy of the variance estimates, we conducted bootstrapping procedures. In bootstrap resampling, random samples with replacement are repeatedly drawn from the original sample. Specifically, we employed the so-called boot-*p* procedure, in which we randomly drew 50 samples (with $n_p = 30$ and $n_p = 40$) with replacements; it has been argued that at least 50 replications are needed to adequately estimate standard errors (Efron and Tibshirani 1986). The means of the bootstrapping parameters can be calculated as follows:

$$\hat{\theta}_B = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b ,$$

where $\hat{\theta}_B$ reflects the bootstrap estimate of some parameter θ , which is based on B number of bootstrap samples; and $\hat{\theta}_b$ indicates a parameter estimate of one particular bootstrap sample b . In addition, the following equation can be used to calculate the standard deviation of $\hat{\theta}_b$:

$$\hat{\sigma}(\hat{\theta}_b) = \sqrt{\frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b - \hat{\theta}_B)^2} .$$

The results in Table 4.6 show that the findings for bootstrapping based on $n_p = 30$ and $n_p = 40$ are similar to the original estimates of the variance components in Table 4.5 and confirm that online and mail surveys produce similar variance components across estimation procedures. To account for the stability of the variance component estimates, we also calculated the standard

errors of the variance components (Brennan 2001; Searle, Casella, and McCulloch 1992), which all were lower than .62, clearly indicating the stability of the estimated variance components.

4.4.3 Optimization of Measurement Designs

The purpose of the G-theory approach is to determine the most cost-efficient design in online and mail surveys that produces a desirable G-coefficient. Nunnally (1978) suggests as a rule of thumb that a G-coefficient of .90 should be the absolute minimum reliability for any applied study, and .95 is the desired level of reliability. As we mentioned previously, we have defined the cost functions C1 for mail and C2 for online surveys. The structure of the cost functions enables us to minimize C_T across n_{persons} , n_{items} , and n_{aspects} , subject to the following constraints: (1) $E\hat{\rho}^2 \geq g$, the desired G-coefficient, and (2) $n_{\text{countries}}$, $n_{\text{business units}}$, n_{items} , n_{persons} , and n_{aspects} are integer values. In this study, we added the following further constraints: $n_{\text{countries}} = 2$; $n_{\text{business units}} = 2$; $n_{\text{persons}} \geq 10$; $n_{\text{aspects}} = 6$; $n_{\text{items}} \geq 3$; and $n_{\text{items}} \leq 10$. If we take into account the assumptions of this design, the cost function, and the suggested constraints, we can minimize C_T over n_{persons} , n_{aspects} , and n_{items} . Tables 4.7 and 4.8 compare the costs of conducting an online or a mail survey against a set of G-coefficients ranging from .80 to .95. In general, the results reveal a significant increase in costs to achieve a G-coefficient greater than .90. Moreover, it is not possible to estimate designs for the .94 and .95 G-coefficient for the business \times aspect interaction in the mail survey. The mail survey requires more observations to achieve an acceptable level of generalizability for comparing countries on different service quality aspects than does the online survey.³

³ We conducted two additional analyses using (1) an equal number of subjects per business unit per country and (2) an equal number of items per aspect. Overall, the findings of these additional analyses are consistent with the findings of the basic analyses.

TABLE 4.6: Variance Components and G-Coefficients Based on Bootstrapping Procedures

Source of Variation	Means Original estimates		Means Bootstrap ($n_b = 30$)		Means Bootstrap ($n_b = 40$)	
	Mail Survey	Online Survey	Mail Survey	Online Survey	Mail Survey	Online Survey
Countries	0.02802	0.17817	0.04279	0.14811	0.03580	0.14877
Business units	0.00496	0.00000	0.00688	0.00158	0.00547	0.00000
Persons	1.72712	1.52264	1.72752	1.58893	1.71607	1.58373
Aspects	0.24220	0.25568	0.22102	0.24027	0.22167	0.22733
Items	0.00988	0.03978	0.01407	0.04735	0.01531	0.04861
Countries \times Business units	0.00000	0.15684	0.01563	0.13305	0.01217	0.14499
Countries \times Aspects	0.09872	0.12375	0.10472	0.12487	0.10962	0.13235
Countries \times Items	0.00000	0.00000	0.00247	0.00055	0.00029	0.00177
Countries \times Business units \times Aspects	0.00000	0.00000	0.02334	0.01155	0.01512	0.00959
Countries \times Business units \times Items	0.01235	0.00000	0.01247	0.00195	0.01211	0.00000
Business units \times Aspects	0.01548	0.04327	0.02030	0.04567	0.01743	0.04302
Business units \times Items	0.00586	0.02283	0.01008	0.02500	0.00909	0.02180
Persons \times Aspects	1.23198	0.88856	1.25999	0.85815	1.25804	0.85216
Persons \times Items	0.00000	0.00000	0.33768	0.49808	0.38226	0.43230
<hr/>						
	Estimated Standard Errors		Estimated Standard Errors Bootstrap ($n_b = 30$)		Estimated Standard Errors Bootstrap ($n_b = 40$)	
	Mail Survey	Online Survey	Mail Survey	Online Survey	Mail Survey	Online Survey
Countries	0.05684	0.08546	0.05684	0.08546	0.03689	0.07178
Business units	0.01407	0.00815	0.01407	0.00815	0.00892	0.00000
Persons	0.18508	0.17138	0.18508	0.17138	0.14324	0.09815
Aspects	0.06536	0.10173	0.06536	0.10173	0.05627	0.07669
Items	0.01268	0.02197	0.01268	0.02197	0.01288	0.01766
Countries \times Business units	0.03138	0.08382	0.03138	0.08382	0.02934	0.08307
Countries \times Aspects	0.06381	0.04378	0.06381	0.04378	0.04708	0.05057
Countries \times Items	0.00771	0.00260	0.00771	0.00260	0.00152	0.00709
Countries \times Business units \times Aspects	0.04269	0.01678	0.04269	0.01678	0.02112	0.02056
Countries \times Business units \times Items	0.01287	0.00654	0.01287	0.00654	0.00996	0.00000
Business units \times Aspects	0.01939	0.03849	0.01939	0.03849	0.01687	0.02403
Business units \times Items	0.00861	0.01813	0.00861	0.01813	0.00755	0.01495
Persons \times Aspects	0.31213	0.21320	0.31213	0.21320	0.21310	0.15283
Persons \times Items	0.19127	0.24104	0.19127	0.24104	0.62401	0.61247

TABLE 4.7: Optimal Research Designs for Comparing Countries Across Aspects

Mail Questionnaire										Online Questionnaire									
Facet Sampling										Facet Sampling									
$E\rho_{c,a}^2$	$\hat{\sigma}^2(\delta_{c,a})$	n_c	n_b	n_p	n_a	n_i	N_t^b	Cost(€)		$E\rho_{c,a}^2$	$\hat{\sigma}^2(\delta_{c,a})$	n_c	n_b	n_p	n_a	n_i	N_t^b	Cost(€)	
.80	.00211	2	2	31	6	5	3720	2953		.80	.00257	2	2	24	6	3	1728	2632	
.85	.00148	2	2	45	6	5	5400	3465		.85	.00187	2	2	33	6	3	2376	2633	
.90	.00093	2	2	71	6	6	10224	4550		.90	.00119	2	2	52	6	3	3744	2634	
.91	.00086	2	2	78	6	6	11232	4819		.91	.00106	2	2	58	6	3	4176	2635	
.92	.00076	2	2	89	6	6	12816	5242		.92	.00095	2	2	65	6	3	4680	2635	
.93	.00067	2	2	104	6	6	14976	5819		.93	.00083	2	2	74	6	3	5328	2636	
.94	.00057	2	2	124	6	6	17856	6588		.94	.00071	2	2	87	6	3	6264	2637	
.95	.00048	2	2	146	6	7	24528	7710		.95	.00060	2	2	103	6	3	7416	2638	

$a \geq g$, the desired G-coefficient.

b Total number of observations = $n_{\text{countries}} \times n_{\text{business units}} \times n_{\text{persons}} \times n_{\text{aspects}} \times n_{\text{items}}$.

TABLE 4.8: Optimal Research Designs for Comparing Business Units Across Aspects

Mail Questionnaire										Online Questionnaire									
Facet Sampling										Facet Sampling									
$E\rho_{bu,a}^2$	$\hat{\sigma}^2(\delta_{bu,a})$	n_c	n_b	n_p	n_a	n_i	N_t^b	Cost(€)		$E\rho_{bu,a}^2$	$\hat{\sigma}^2(\delta_{bu,a})$	n_c	n_b	n_p	n_a	n_i	N_t^b	Cost(€)	
.80	.00033	2	2	245	6	10	58800	13098		.80	.00093	2	2	209	6	3	15048	2647	
.85	.00024	2	2	416	6	10	99840	20969		.85	.00062	2	2	376	6	4	36096	2690	
.90	.00015	2	2	1111	6	10	266640	52962		.90	.00041	2	2	549	6	6	79056	2764	
.91	.00014	2	2	1615	6	10	387600	76162		.91	.00038	2	2	804	6	6	115776	2784	
.92	.00012	2	2	2901	6	10	696240	135359		.92	.00033	2	2	754	6	7	126672	2810	
.93	.00010	2	2	13177	6	10	3162480	608384		.93	.00029	2	2	849	6	8	163008	2848	
.94										.94	.00025	2	2	736	6	10	176640	2899	
.95										.95	.00021	2	2	2266	6	10	543840	3021	

$a \geq g$, the desired G-coefficient.

b Total number of observations = $n_{\text{countries}} \times n_{\text{business units}} \times n_{\text{persons}} \times n_{\text{aspects}} \times n_{\text{items}}$.

4.5 Discussion

This study aims to extend the sparse literature on the comparison of response quality between online and mail surveys by providing evidence about the generalizability of both. The results reveal the reliable response quality of online surveys; their generalizability is comparable to that of mail surveys. These findings are similar to those of recent studies pertaining to the equivalence between online and mail surveys (Cobanoglu, Warde, and Moreo 2001; Epstein et al. 2001; Knapp and Kirk 2003) and indicate that factors that could cause differences, such as computer anxiety, privacy concerns, response processes, or the way people perceive questions onscreen or on paper, are reduced as people become more familiar with the Internet.

In addition to shedding more light on the current debate about whether the response quality of increasingly popular online surveys is comparable to mail surveys, the findings add to existing literature by viewing response quality in terms of the data collection costs, which involves one of the main reasons more researchers are using online surveys. Furthermore, G-theory allows us to pursue two different goals: (1) maximizing response quality given a certain cost level or (2) minimizing costs given a desired response quality level. In a specific example, to reach a G-coefficient of .90, an online survey would cost €2,634 if the country \times aspect interaction is the object of measurement, whereas the corresponding mail survey would cost €4,550. If the G-coefficient is .95, the cost for the mail survey increases considerably to €7,710, whereas the cost for the online survey remains almost constant (€2,638). Because online surveys provide equally generalizable results, the results imply that researchers can minimize data collection costs while achieving their desired or required generalizability levels.

Another key result is that the country \times aspect interactions explain a larger percentage of the variance of the service quality assessment in case of online surveys. Therefore, the SERVQUAL aspects have different meanings in different countries and are thus country specific. If researchers wish to generalize across countries, online surveys would be better suited because more of the variance is explained. This finding demonstrates that the type of research can influence the choice of the research mode.

In addition, we illustrate how G-theory can be used to compare online and mail surveys while taking data collection costs into account and thereby

demonstrate the economic benefits of G-theory. Although the two-stage procedure with the G-study and subsequent D-study may seem time consuming and costly, it optimizes the measurement design. Researchers can determine to what extent the distinct aspects account for variance in the SERVQUAL measure and sample accordingly, ensuring that money is not being wasted on unnecessary data collection.

Optimization is especially interesting for mixed-mode surveys, in which companies use both traditional and new survey modes, such as mail and online surveys, in combination. If a customer database does not contain all potential respondents' e-mail addresses, mixed-mode surveys can enable researchers to contact those members of the sample with a regular mail survey (Dillman 2000). Oftentimes, the mail survey also contains a hyperlink to the online questionnaire to reduce costs while giving respondents the choice of their preferred survey medium.

Researchers also can use a D-study design in which online and mail surveys are combined in a mixed-mode survey and the level of generalizability and data collection costs are calculated relative to the share of the survey mode in the study design. With respect to the four criteria to evaluate a data collection technique, mixed-mode surveys entail several advantages, such as (1) reduced response times due to the faster response to online surveys (Illieva, Baron, and Healey 2002); (2) increased response rates, according to recent evidence that response rates can be maximized through the use of several contact modes (Dillman 2000; Dillman et al. 2001); (3) reduced costs because fewer members have to be contacted via more expensive mail surveys; and (4) generalizable results.

4.6 Limitations and Future Research Guidelines

The findings are subject to several limitations. We examined the generalizability of online and mail surveys in the context of a customer satisfaction survey in the United States and United Kingdom. Additional studies should examine whether the findings hold for different contexts, such as different scales. In addition, research should examine whether the findings hold for other countries, especially if a country has a lower rate of Internet adoption. Furthermore, we decided to focus on online and mail surveys because the type of research for which mail surveys are used is most likely to be supplemented by

online surveys. It could also be interesting to examine how alternative modes, such as telephone surveys, compare to online surveys. Another limitation of this study is that potentially important sources of variance, such as culture, customer characteristics, and different channels, were not incorporated into the analysis. Further research should investigate whether such other sources account for additional variance. Finally, we used G-theory as a framework to examine response quality in relation to data collection costs. Additional research should verify the quality of online surveys by examining different aspects of response quality, such as measurement invariance or response styles.

CHAPTER 5: INDIVIDUAL DIFFERENCES IN MOTIVATION TO PARTICIPATE IN ONLINE SURVEYS AND THE EFFECT ON RESPONSE RATE AND RESPONSE QUALITY

The majority of online research is conducted via online access panels since they promise high response rates, sampling control, access to populations that are hard to reach, and detailed information about respondents. To sustain a critical mass of respondents, to overcome panel attrition, and to recruit new panel members, it is important to understand how one can predict and explain what motivates people to repeatedly participate in online surveys. Therefore, the aim of this study is to enhance the existing survey literature by developing a survey participation inventory that identifies the underlying response motives that stimulate panel members to participate in online surveys. In addition, respondent motives were linked to important response parameters and respondents were assigned to three clusters that could be characterized as voicing assistants, reward seekers, and intrinsics. Important implications concern the trade-off between response rate and response quality.

*The value of marketing decisions is largely dependent
on the worth of the sample data they are based on.
Therefore, quality samples need quality panels!
(Prof. Dr. K. De Wulf, InSites Consulting)*

5.1 Introduction

Online surveys present one of the fastest growing segments in market research (ESOMAR 2004a; Hogg 2002). Most online consumer research is conducted using online access panels (Göritz 2004), which are defined as a pool of people who have agreed to participate in online surveys on a regular basis (Göritz, Reinhold, and Batinic 2002). The main advantages of online access panels are reduced costs associated with locating appropriate respondents and respondents' immediate availability. Additional benefits are increased response rates, shorter field times, and easy identification of key sample segments (Göritz 2004). At the same time, unavailability of representative online sampling lists and privacy regulations concerning unsolicited e-mails (<http://www.informationcommissioner.gov.uk/eventual.aspx?id=35>) push researchers to use online access panels. An effective online research panel requires a critical mass of members that is motivated to repeatedly participate in online surveys. In addition, uneven participation and panel attrition force panel managers to constantly encourage contribution and recruit new panel members (Winer 1983). Therefore, it is critical for researchers to obtain relevant information on how to predict and explain what motivates panel members to participate in online surveys. To improve the effectiveness of online access panels, researchers should also have a better understanding of whether differences in motivation have an impact on response quality parameters. Yet, research on response motives has not been the main focus of investigation. Considering the contemporary trend and future prospect of online access panels, the purpose of the current study is to extend research on the effectiveness of online panels by examining a number of important issues that have been left unexplored so far.

First, the existing survey literature has had a strong focus on incentives, an extrinsic motive, as the most effective response stimuli (e.g., Church 1993; Kanuk and Berenson 1975; Yu and Cooper 1983). Yet, concentrating on incentives only offers limited theoretical understanding of the psychological processes that determine why people repeatedly participate in online surveys.

Psychological theory stresses the importance of intrinsic motivation to explain behavior (e.g., Hunt 1965). More specifically, intrinsic motives such as topic interest or curiosity have been advanced as important response motives (e.g., Heberlein and Baumgartner 1978; Sheehan and McMillan 1999). In addition, it has been suggested that particularly in an online panel context social norms are important drivers of survey participation (Groves, Singer, and Corning 2000). Online access panels present a social context in which researchers and participants interact with each other. Some online access panels are even organized in the form of online virtual communities with discussion forums, everyday quick polls, announcement of prize winners, and latest news (e.g., <http://www.opinionworld.co.uk/>). In this context, respondents feel morally obliged to participate and help researchers and receive satisfaction from the fulfillment of their civic duties. Thus, the social nature of motives is an important element when considering participation in online access panels. The current study aims to enhance the online survey literature by developing a survey participation inventory that systematically investigates a complete set of response motives in online access panels along two dimensions: extrinsic versus intrinsic motivation and self versus other orientation.

Second, respondent motives to participate in online access panels should be linked to important response parameters, since researchers might have to weigh up high response rates with a potential loss in response quality. For example, respondents in online access panels could be highly intrinsically motivated or participate only because of the money they can win (e.g. "The free Get Paid to Take Surveys Online Guide;" <http://www.surveys4money.com/>). Even though both motives might be associated with high response rates, they also could lead to response styles such as yea or nay-saying, extreme responses, or midpoint responding (Baumgartner and Steenkamp 2001; Greenleaf 1992a, 1992b; Shulman 1973). For example, respondents who join an online access panel because of the incentives offered might engage in satisficing and only use midpoint responding. In contrast, intrinsically motivated respondents might be more enthusiastic and positive in their answers. Thus, it is important to investigate the relationship between response motives and response quality parameters.

A final important, but so far unexplored, issue relates to the possible existence of respondent clusters that are driven by different underlying motives

to participate, e.g. intrinsic versus extrinsic, or high versus low levels of motivation. Knowledge on different motivational clusters is insightful for panel managers in terms of sample composition. Besides demographic profiles of respondents, underlying response motives might factor into weighting or adjustment strategies that are frequently used in online access panels. Therefore, there is a need to examine whether researchers have to account for different respondent segments in online access panels.

To address the three aforementioned issues, this study seeks to extend the existing literature by investigating which motives drive members of online access panels, whether the different motives have a differential impact on response likelihood and quality, and whether respondents can be grouped into different respondent clusters. We begin by developing a survey participation inventory (SPI), which identifies and measures the importance of several response motives. We then use the SPI in a series of subsequent studies to identify the relationship between the motives and response rate and response quality and to cluster respondents in online access panels.

5.2 Conceptual Development of the Survey Participation Inventory (SPI)

The primary focus of the survey literature has been on the effect of design factors on response rates and it has been repeatedly ascertained that incentives are most effective in increasing response rates (e.g., Church 1993; Dillman 2000; Fox, Crask, and Kim 1988; Heberlein and Baumgartner 1978; Kanuk and Berenson 1975; Yammarino, Skinner, and Childers 1991; Yu and Cooper 1983). While impressive in scope, this literature stream offers limited theoretical understanding of what motivates panel members to participate regularly in online surveys (Goyder 1987). Beyond the design considerations of incentives, the motivation literature states that such external rewards stimulate extrinsic motivation (Ryan and Deci 2000). Extrinsic motivation implies that the reason for participating in a survey is something other than an interest in the survey itself, such as reward or recognition, which motivates respondents to participate (Deci and Ryan 1985). In contrast, intrinsic motivation implies that people respond for no reward other than the interest and enjoyment that accompanies the activity (Deci and Ryan 1985). Contemporary views of intrinsic motivation include both affective and cognitive elements. While the cognitive evaluation theory of Deci and Ryan (Deci and Ryan 1985), for example, stresses self-determination and

competence as important aspects of intrinsic motivation, other theorists have proposed affective components such as interest (Izard 1977), fun (Pretty and Seligman 1983), and flow (Csikszentmihalyi 1975). Similarly, evidence from the survey literature suggests that intrinsic response motives such as topic interest or curiosity are important response motives (e.g., Heberlein and Baumgartner 1978; Sheehan and McMillan 1999).

In addition to intrinsic and extrinsic motivation, social norms play an important role in online panels that are often organized in the form of online virtual communities. Daugherty et al. (2005) confirm that online access panels are able to evoke a sense of community. Sustained and repetitive interaction through research invitations, discussion forums, and polls, create a sense of belonging and participants feel rewarded because they have personally helped to generate knowledge. Groves, Singer, and Corning (2000) endorse that social norms play an important role in the social context of online access panels. Whereas some respondents participate based on more self-centered reasons such as incentives where they seek financial benefits for themselves, other respondents might feel morally obliged to participate and help researchers and receive satisfaction from the fulfillment of their civic duties.

Taken together, we can isolate two dimensions for examination: (1) the classification into intrinsic and extrinsic response motives; and (2) the degree of socially shaped norms, classified as self or other orientation. In the next paragraph, we systematically review and integrate evidence from the existing survey literature to develop a comprehensive set of response motives along these two dimensions.

Extrinsic - Self

As mentioned earlier, the survey design literature has identified incentives as one of the most effective response stimuli. According to social exchange theory, participation in surveys is a function of perceived costs, expected rewards, and trust (Dillman 2000). In the context of online access panels, incentives are purely aimed at increasing the reward dimension of the reward, cost, and trust matrix. Singer and Kulka (2000) state that the rationale for using incentives is that people who are otherwise not motivated to participate are stimulated by the monetary inducement. Likewise, several authors confirm that incentives are most effective in the absence of other, intrinsic stimuli, which

substantiates that a subgroup of respondents participates because of the monetary incentive (Groves, Singer, and Corning 2000; McDaniel and Rao 1980; Nederhof 1983). Thus, respondents who are mainly motivated by incentives are extrinsically motivated and driven by a desire to seek financial benefits for themselves.

Extrinsic - Other

Besides incentives, non-monetary returns such as recognition are frequently mentioned as factors providing rewards to respondents. Dillman (2000), for example, states that being considered positively by others and receiving positive regard provides value to many people. Similarly, Warwick and Lininger (1975) and Krosnick (1999) mention that gratification from the successful performance of the respondent's role motivates respondents to participate. In addition, obligation has been proposed as an extrinsic motive that stimulates respondents to participate. While some researchers consider obligation as a synonym for helping and therefore deem it as a form of intrinsic motivation, the survey literature regards it as a form of introjected regulation. This means that respondents do not participate freely but act instead to avoid feeling guilty for not doing it (Alexandris, Tsorbatzoudis, and Grouios 2002). Thus, respondents who agreed to join an online access panel regulate their behavior and regard it as their duty to participate in subsequent survey requests. Bosnjak, Tuten and Wittmann (2005) ascertain that moral obligation plays an important role in predicting participation intentions in online surveys. Similarly, Groves, Cialdini, and Couper (1992) as well as Albaum, Evangelista and Medina (1998) state that some people feel obliged to respond and Krosnick (1999) mentions social responsibility as a driver of survey participation. In contrast to incentives, respondents who are motivated by a need for recognition or a sense of obligation focus less on their own financial benefits than on value derived from the interaction with the researcher and the adherence to social norms. Therefore, need for recognition and obligation can be classified as extrinsic – other motives.

Intrinsic - Self

Intrinsic motives such as enjoyment, curiosity, or topic interest have been presented as important drivers of survey response. Several researchers provide evidence that when respondents perceive a topic as salient, they are more likely

to return the questionnaire (Heberlein and Baumgartner 1978; Helgeson, Voss, and Terpening 2002; Roberson and Sundstrom 1990; Sheehan and McMillan 1999; Van Kenhove, Wijnen, and De Wulf 2002). When the respondent is interested in the subject of the study, the cost of responding may be reduced, and personal input to the study may be judged as more important (Heberlein and Baumgartner 1978). Evangelista, Albaum, and Poon (1999) also find that topic interest was the most important self-reported reason for participating in surveys. Closely related to interest, several researchers have proposed curiosity as a main driver of survey participation. Some respondents are just curious what they are contacted for and what the research is about. Porst and van Briel (1995) support that 14.9% of their respondents indicated interest and curiosity as their main stimulus for panel participation. In addition, Dillman (2000) states that some people just enjoy answering questionnaires regardless of content. Also, Porst and van Briel (1995) find that 9% report enjoyment as their main driver of participating in a panel. While this number seems rather low, several authors state that enjoyment is greater for online surveys (Knapp and Kirk 2003; McDonald and Adam 2003). Forrest (1999), for example, reports that in comparison to mail surveys, respondents to online surveys seem to enjoy the process more, probably due to novelty of the research medium, inclusion of color, graphics and sound, and 24/7 access to the survey, which allows respondents to answer the questionnaire at their convenience (Forrest 1999; McDonald and Adam 2003). Interest, enjoyment, and curiosity are all motives that present an inherent intrinsic satisfaction for respondents.

Intrinsic - Other

An obvious motivator for respondents to participate is that they like to voice their opinion on a variety of matters. Respondents who are driven by this type of motivation feel honored and believe that they can have an impact when they voice their opinion. This type of motivation is often triggered by including statements such "having your opinion count" or "make your voice heard" into the survey invitation (Dillman 2000; Groves, Cialdini, and Couper 1992). Evangelista, Albaum and Poon (1999) find that 17.1% of respondents in Australia are mainly motivated by the fact that their opinion counts. In addition, several authors suggest that respondents are driven by a feeling of altruism or helping to respond to a survey request (Groves, Cialdini, and Couper 1992; Krosnick 1999;

Warwick and Lininger 1975). Even though Linsky (1975) concludes from his meta-analysis that altruism does not serve as a motivational base for inducing survey participation, Groves, Cialdini, and Couper (1992) ascertain that including a helping request increases response rates significantly. Also, Dillman (2000) and Porst and von Briel (1995) list asking for help as a reason for participation. Based on a content analysis of open-ended responses of 140 participants in a German panel study, Porst and von Briel (1995) ascertain that 31% are driven by altruistic reasons. Similarly, Evangelista, Albaum, and Poon (1999) find that helping was the second most important self-reported reason for participating in surveys in Australia and Hong Kong. Respondents that are motivated by helping or who like to give their opinion derive value from the feeling that they could do something for someone else. Thus, giving an opinion and helping are both motives that enhance or maintain a person's self-concept and can therefore be classified as intrinsic motivation, but one which derives from the interaction with others.

All these factors have been identified in separate studies in a general way with incentives being the most dominant. However, no previous research has identified and tested a more complete set of individual differences that could drive response. Furthermore, while many of these factors have been mentioned, they have not been concurrently investigated in a large empirical study. Therefore, the aim of this study is to develop the survey participation inventory, as stated in our first research question (see Table 5.1):

How can the Survey Participation Inventory, consisting of extrinsic-intrinsic motivation and self-other dimensions be empirically substantiated?

TABLE 5.1: Typology of Response Motives

Orientation	Intrinsic Motivation	Extrinsic Motivation
Self	Interest Enjoyment Curiosity	Incentives
Other	Helping Give Opinion	Obligation Need for recognition

5.2.1 Survey Participation Inventory and Response Parameters

The crucial question for researchers is not so much the type of motivation *per se*, but whether the underlying motives have an effect on important response parameters such as response rate and response quality. Response rate has received a great deal of attention in the survey literature since it has been used as the main criterion for judging the desirability of design factors (e.g., Kanuk and Berenson 1975; Yu and Cooper 1983). Response rates are considered important since it has been argued that the best way to reduce nonresponse error and thus secure representative results is by maximizing response rates (Hansen 1980). Yet, focusing on response rates only neglects the possibility that design factors or response motives may yield an unrepresentative sample increment or influence respondents' answers to questions. Therefore, response quality has been proposed as an additional criterion (e.g., Goetz, Tyler, and Lomax Cook 1984; 1978). Response quality is an umbrella term and normally incorporates different measures for the accuracy and completeness of respondent answers, depending on characteristics of the questionnaire. Completeness, if possible, is measured by looking at open-ended questions, "don't know" responses, or missing answers (Goetz, Tyler, and Lomax Cook 1984). Accuracy refers to a form of response bias, also called response style, which is defined as the tendency "to respond systematically to questionnaire items on some basis other than what the items were specifically designed to measure" (Paulhus 1991, p. 17). Specifically, Greenleaf (1992a) and Baumgartner and Steenkamp (2001) compile a list of six potential response styles: yea-saying, nay-saying, extreme responding, midpoint responding, means and standard deviations. Examining the impact of response motives on both response rate and response quality is extremely important in the context of online access panels since different response motives might be positively related to response rates while having an impact on response styles.

The previous paragraph on the development of the response typology already cites empirical evidence that establishes a relationship between the response motives as listed in the four quadrants of the typology and response rate. Yet, existing evidence on the relationship between response motives and response quality is scarce or inconclusive (e.g., Hansen 1980; McDaniel and Rao 1980). Considering incentives, for example, some researchers have argued in line with social exchange theory that incentives improve response quality, as

they increase the attention respondents devote to the task for which they are being paid and thus answer more carefully and more completely (Goetz, Tyler, and Lomax Cook 1984; McDaniel and Rao 1980). Importantly, this evidence was derived from surveys in which prepaid monetary incentives were used. According to Dillman (2000) such incentives are not considered an external reward, but instead are viewed as a social gesture. As such, the psychological value of the incentive would elicit a feeling of reciprocity (Gouldner 1960) and thus intrinsically motivate the respondent to participate in the survey. The incentive structure in online surveys and online access panels in particular is very different from traditional survey modes. Respondents either take part in a lottery or the panel uses some form of point system. In either case, the incentives are received post-hoc, and, in the case of lotteries, not everyone is rewarded for participation. Therefore, these incentive schemes clearly represent a payment for the service provided by the respondent. Following cognitive evaluation theory, receiving such an extrinsic reward should undermine the effort invested into the task and the performance (Deci and Ryan 1985). Evidence from the mail survey literature has verified that, according to self-perception theory, incentives reduce response quality (Hansen 1980). Self-perception theory basically stipulates that if respondents perceive their behavior to be elicited by external factors such as incentives, they discount internal motives as the cause of their behavior. Therefore, respondents feel less commitment (Hansen 1980) and tend to satisfice (Krosnick 1999) or use other response styles such as nay-saying (Baumgartner and Steenkamp 2001).

Since we have exposed a new typology of response motives and such a detailed investigation of differences in response motives presents new ground, little literature is available that has empirically addressed the relationship between obligation and recognition and response quality. Need for recognition is likely to work as a form of controlling feedback, where respondents are rewarded by the recognition of the researcher when they put effort into filling in the questionnaire properly. Such an external orientation could implicate response styles that are driven by a sense of social desirability. With respect to obligation, norm theory suggests that people who feel morally obliged to return the questionnaire should also feel obliged to put in effort and complete the questionnaire. Yet, if the sense of duty dominates and is unaligned with other

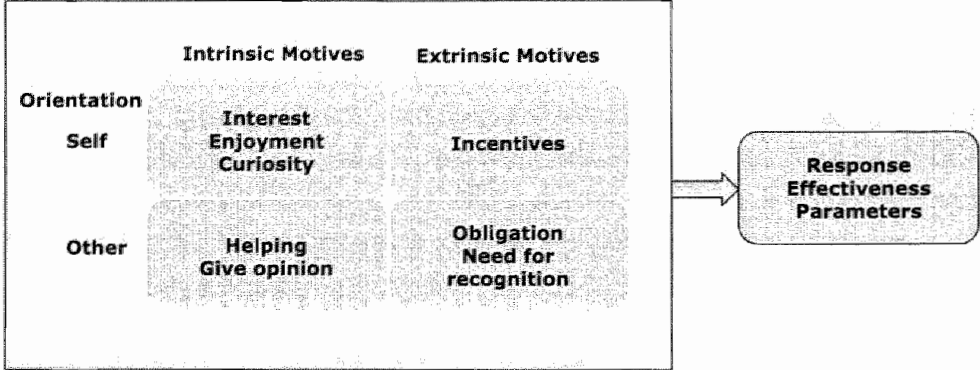
motives, it could be that respondents are more critical and negative in order to create a balance between give and take.

Very little empirical research has examined the relationship between intrinsic motivation and response quality. Some authors suggest that topic interest is positively related to effort. Groves, Presser and Dipko (2004) argue that when respondents are interested and involved in the topic of the survey, they engage in extensive cognitive processing, similar to the central route of persuasion (Petty and Cacioppo 1986). Respondents who are mainly driven by topic interest may think through their answers more carefully and more elaborately (Bickart and Schmittlein 1999). A similar line of reasoning can be applied to enjoyment and curiosity, where respondents who are really curious, or enjoy the process of responding are likely to put in more effort to question interpretation, retrieval, judgment, and response selection, a response approach that Krosnick calls optimizing (1991). Since respondents who are driven by helping motives want to assist the researcher in collecting data, it is likely that they devote considerable effort to answering the questionnaire. Likewise, respondents who like to give their opinion should be very concerned about communicating their view clearly and completely to the surveyor. While high levels of effort are positive since they are traditionally associated with high performance (Krosnick 1999), high effort in the context of survey participation could also bring about response styles such as more positive and extreme responding.

Since little empirical research is available that has empirically addressed the relationship between a detailed typology of response motives and response quality parameters, the second research question is (see Figure 5.1):

How is the typology of response motives for online access panels related to effectiveness parameters such as response rate and response quality?

Figure 5.1: Conceptual Framework



5.2.2 Clustering Respondents based on the Survey Participation Inventory

Given the variety of response motives, it seems valuable to also investigate potential differences among groups of respondents. Experience with online panels shows that there is a group of respondents which is always participating. It seems that die-hard participants have a different predisposition and thus motivational orientation towards survey participation. This is supported by evidence from the mail survey literature, which verifies that refusers demonstrate quite different attitudes and behaviors than nonrespondents and that such behavior remains quite consistent across surveys (Brennan and Hoek 1992). Yet, we know little about the specific set of motives that distinguish certain types of respondents. A cluster analysis could reveal different respondent segments based on the type and level of motivation and the corresponding effect on response quality parameters. This would imply, for example, that even though a sample is representative in terms of demographics, an overrepresentation of highly motivated respondents would still affect the results. Therefore, it is critical for panel managers to obtain relevant information on potential clusters. The more they know, the better could differences in motives be used in the set-up of weighting or adjustment techniques. Therefore, the third research question of this study is:

Can respondents be grouped into respondent clusters that are driven by different underlying motives to participate?

5.3 Research Design and Methodology

To empirically validate the typology, examine the effect of different motives on response quality parameters, and to cluster respondents, we divided the research section into two parts. The first section discusses the development of the survey participation inventory (SPI), which measures the response motives as identified in the typology. One qualitative and two quantitative pretests were conducted to provide empirical validation of this measurement instrument. In section two, the empirical study is described in which we administered the SPI to members of an online access panel. The aim of this study is to identify the impact that the different response motives have on response quality parameters and whether respondents can be grouped into clusters that are driven by different underlying motives to participate.

5.3.1 *The Survey Participation Inventory (SPI)*

Unfortunately, no existing scales from the survey literature were available to measure the response motives that we identified in the typology. Therefore, we first developed the SPI to be able to identify individual differences in why people participate in online panels and the corresponding response quality.

In a first step, we conducted 14 in-depth interviews with members of an online panel. The results of these interviews were content analyzed and found to confirm the list of 8 motives as developed in the typology. With respect to intrinsic reasons relating to self, respondents said that they are participating because they find it interesting (8 respondents), they just enjoy it (3 respondents), and they are curious (2 respondents). With respect to intrinsic other-oriented motives, respondents said that they like to give their opinion (5 respondents) and they like to help (2 respondents). Extrinsic, self-oriented reasons were the rewards that they would get upon completion (4 respondents), while the recognition they would get from the researcher (1 respondent), and feelings of obligation to participate (2 respondents) were mentioned as other-oriented motives. Some excerpts of what respondents said are presented in Table 5.2.

Table 5.2: Excerpts from Qualitative Interviews with Members of an Online Panel

Motives	Excerpts
Interest	"...Seems interesting..." "...I am interested in marketing..." "...I find the topics interesting..." "...I am interested in the topics..." "...The topics are interesting..." "...I am just interested..." "...I am interested in what happens around me..." "...I find online surveys nice and interesting..."
Enjoyment	"...I find it enjoyable..." "...I just find it pleasurable..." "...I just find it fun to participate..."
Curiosity	"...I am curious; I want to know what's at hand..." "...I have to admit, I am also a bit curious about the latest trends in marketing..."
Give Opinion	"...I appreciate that I can give my opinion..." "...I like that I can give my opinion on a variety of matters..." "...I like to give my ideas and opinions..." "...It is important to me that my opinion counts..." "...I like to give my opinion..."
Helping	"...From University, I know how difficult it is to conduct marketing research. Therefore, I decided to help by participating..." "...I am supporting a good cause..."
Incentives	"...Besides the fact that I can give my opinion and I find it interesting, I also get points for participation. These points are not the most important reason for my participation, but they are a nice supplementary bonus..." "...I though it would be nice; that had also to do with the point system..." "...Sometimes, you can win something..." "...I participate because I can get something for the points that I collect for participation..."
Need for recognition	"...I feel flattered that they are interested in my opinion..."
Obligation	"...I feel obliged to participate. If I don't do it, no one does..." "...I was asked to participate, so I just did..."

Together with 2 market researchers and based on our literature review, we developed an initial set of 50 questions. Most questions were based on scales in similar domains and were adapted to this specific research context. The questions on interest were mainly based on Pierce et al. (2003). Enjoyment was asked with questions derived from the Intrinsic Motivation Inventory (IMI) as well as Amabile et al. (1994). The helping questions were based on existing altruism scales (e.g., Bettencourt and Brown 1997; Posdakoff and Mackenzie 1994; Yoon and Suh 2003). Curiosity linked to the challenge dimension in Amabile et al. (1994). The questions on giving an opinion were loosely based on the civic virtue/engagement literature (Bettencourt and Brown 1997; Kirlin 2002; Posdakoff and Mackenzie 1994), but were adapted to this specific situation. The questions on incentives were adapted from the compensation dimension in Amabile et al. (1994). Obligation was measured with questions from Diamond

and Kashyap (1997) as well as de Ruyter and Wetzels (2000). The outward dimension of Amabile et al. (1994) was used for the questions on recognition. All items were formatted as seven-point Likert scales that ranged from "strongly disagree" (1) to "strongly agree" (7). The content validity of the items was assessed by 3 market researchers and 2 academic experts. The five judges had to decide whether each item was representative/not representative for the respective construct. The results showed 400 pairwise agreements out of a total of 500 possible (interjudge agreement of 0.80). Thus, the PRL reliability measure (Rust and Cooil 1994) was 0.98. Therefore, we can be fairly confident about our judges' classification. Next, the SPI was validated empirically by conducting two pretests that are described in the following paragraphs.

Pretest 1

To ascertain reliability and validity of the SPI, the 50-item questionnaire was tested quantitatively with 149 undergraduate students, of which 41.6% were male and 58.4% were female. All students were entered into a lottery draw where they could win one of 25 vouchers of 10€.

Purifying the measure

First, we checked the item-total correlation per construct. Six items with an item-total correlation of less than 0.35 were deleted. Next, internal consistency of the constructs was examined with Cronbach's alpha, which ranged from 0.640 to 0.840 (interest 0.809; curiosity 0.847; enjoyment 0.840; give opinion 0.755; helping 0.810; recognition 0.694; incentives 0.795; obligation 0.640). Three additional items were removed to increase Cronbach's alpha (interest 0.822; curiosity 0.852; enjoyment 0.844; give opinion 0.772; helping 0.851; recognition 0.741; obligation 0.660). Next, the data was subjected to exploratory factor analysis. Specifically, principal axis factoring with Oblimin rotation as implemented in SPSS release 13 was employed. This resulted in a 13 factor solution using the Kaiser criterion. According to Churchill (1979), exploratory factor analyses during the purification steps tend to produce more dimensions than can conceptually be identified. Some items, for example, did not share the common core of the construct and produced additional dimensions. Yet, the eight factors as identified in the typology could clearly be identified.

Pretest 2

In order to investigate discriminant validity, we administered the reduced 32-item SPI to members of an online University panel where we included measures that could be considered closely related to intrinsic and extrinsic motivation: the aspiration index (measuring extrinsic and intrinsic goals) (<http://www.psych.rochester.edu/SDT/measures/aspir.html>), need for cognition (Areni, Ferrell, and Wilcox 1999; Cacioppo, Petty, and Kao 1984), and a self-reported level of effort in filling out this online survey (based on the Intrinsic Motivation Inventory (<http://www.psych.rochester.edu/SDT/measures/intrins.html>)). In order to examine whether the questions truly measured the SPI rather than the tendency to report in a socially desirable way, we also administered a short version of the social desirability scale (Crowne and Marlowe 1960).

We received 111 usable responses (response rate 16%), of which 48.6% were male and 51.4% were female. The majority of respondents was between 19 and 35 years old (19-25: 65.2%; 26-35: 26.8%; 36-45: 1.8%; 46-55: 2.7%; 56-65: 1.8%; >65:0.9%). Most respondents were students (73.2%), 17.9% were employed.

All responses were subjected to confirmatory factor analysis in order to investigate the psychometric properties. Cronbach's alpha ranged from 0.742 to 0.895 (interest 0.779; give opinion 0.860; enjoyment 0.742; curiosity 0.895; helping 0.824; recognition 0.819; incentives 0.883; obligation 0.809). Again, we found significant deviation from multivariate normality (Mardia 1970; Srivastava 1984) (all tests $p < 0.001$) and therefore used the Satorra-Bentler-scaled χ^2 statistic (χ^2_{SB}) with robust standard errors (Satorra and Bentler 1994). Confirmatory factor analysis for the 8-factor solution provided an acceptable model fit (χ^2_{SB} (436) = 632.118; RMSEA = 0.064; NFI = 0.902; NNFI = 0.963; CFI = 0.967; IFI = 0.968; RFI = 0.889).

Next, we separately assessed the discriminant validity between the SPI, modeled as a second order construct, and the aspiration index, need for cognition, and effort. Discriminant validity was assessed by constraining the estimated correlation parameter (Φ_{ij}) between each pair to one and then performing a Satorra-Bentler χ^2_{SB} difference test (Satorra and Bentler 2001) on the values obtained for the constrained and unconstrained models (Anderson and Gerbing 1988). In all cases, the unconstrained model showed a superior fit to the

constrained model (need for cognition: $\Delta \chi^2_{SB}(1) = 45.28, p < 0.001$; aspiration index: $\Delta \chi^2_{SB}(1) = 7.00, p = 0.008$; effort: $\Delta \chi^2_{SB}(1) = 14.34, p < 0.001$). As all chi-square differences were significant ($p < 0.001$), we concluded that the SPI exhibits discriminant validity.

To demonstrate that the SPI is not affected by a social desirability bias, we split the data into two groups using a median-split procedure. Group 1 contained respondents that scored low on the social-desirability scale, group 2 comprised respondents who reported a high level of social desirability. Due to the small sample size, we compared the variance-covariance matrices of the two groups using the GLM procedure as implemented in SPSS release 13. Box's test of equality of covariance matrices confirmed that the observed covariance matrices of the SPI were equal across groups (Box M = 40.728; $F(36, 39743) = 1.044; p = 0.397$). In addition, multiple comparison tests revealed that the means of the SPI did not differ across the high and low group since the Bonferroni-type adjusted significance level was exceeded for all variables. Thus, we concluded that the SPI is not affected by social desirability.

5.3.2 Empirical Study

The aim of this study was threefold. First, the SPI should be empirically validated in a real online access panel. Second, the impact of the different response motives on response rate and response quality should be examined. Third, we were interested whether respondents could be grouped into several clusters that are driven by different underlying motives to participate, e.g., intrinsic versus extrinsic, or high versus low levels of motivation.

The sample was recruited from one of the largest online panels in Belgium. The procedure was as follows: First, respondents received an e-mail invitation with a request to participate in a regular market survey. A hyperlink directed them to the Web questionnaire. Double entries could be avoided by using a unique URL for each respondent. The initial, regular market survey was used later on to determine the response quality per respondent. To avoid any effects due to the content of the questionnaire, thirteen different surveys were used that differed significantly in their topic. They ranged from the evaluation of mobile phones, beer, cars, a variety of fast moving consumer goods, to advertising

campaigns, medical foot cream, health, food supplements and public transport. Topic salience was controlled with a scale from Srinivasan and Ratchford (1991).

After they completed the initial questionnaire, respondents were asked to fill-in the 32-item SPI and to provide a self-assessed level of effort and response quality in filling in the initial market survey. In addition, respondents had to divide 100 points over the different response motives (see appendix A).

In total, 43,841 respondents participated in one of the thirteen initial, regular market surveys. These market surveys had an average response rate of 33%. From the 43,841 participants, 6036 also filled in the SPI. This resulted in a response rate of 14%.

Analyses

The two main dependent variables are response likelihood and response quality. The response likelihood per respondent was assessed by counting (1) the number of online surveys the respondent was invited to participate in in 2005, (2) how many of these the respondent filled in.

This information allowed us to compute a measure of response likelihood: *Response rate* was calculated as the number of questionnaires the respondent filled in, divided by the number of invitations, multiplied by 100.

Next, the response quality of every respondent had to be estimated in order to answer the second research question. Recall that respondents participated in a regular online market survey before they filled in the SPI. To measure response quality, we first of all asked respondents about their self-assessment of effort and response quality. In addition, we created an objective estimate of response quality by analyzing the accuracy of respondent answers (Goetz, Tyler, and Lomax Cook 1984; Hansen 1980; McDaniel and Rao 1980). Bias or inaccuracy was assessed by calculating several indicators of response styles. In line with Baumgartner and Steenkamp (2001), response style was defined as the tendency to systematically respond to survey questions irrespective of the content. Baumgartner and Steenkamp (2001) and Greenleaf (1992b) provided an overview of different measures of response style, which we assessed based on Weijters et al. (2004).

- *Acquiescence response style (ARS)*: Every time a respondent marked "strongly agree", s/he received two points, every time s/he marked

"agree", s/he received one point. All counts were added to derive an index for ARS.

- *Disacquiescence response style (DARS)*: DARS basically is the reverse of ARS where every time a respondent marked "strongly disagree", s/he received two points, every time s/he marked "disagree", s/he received one point. All counts were added to derive an index for DARS.
- *Extreme response style (ERS)*: To assess ERS, we computed how often respondents marked the extreme answer categories (strongly agree/strongly disagree).
- *Midpoint responding (MPR)*: MPR was assessed by counting the number of times a respondent marked the neutral midpoint of the scale.
- *Intra-subject mean (ISM)*: ISM was defined as the mean of all answers and measured the general tendency to agree or disagree with statements and questions independent of specific item content.
- *Intra-subject standard deviation (ISSD)*: ISSD was the standard deviation of all answers and assessed the tendency to use a wide (or narrow) range of response intervals around the mean response.

Because of the different length and content of the thirteen initial market surveys, the calculation basis for response quality differed for every survey. Therefore, we standardized all measures for response quality.

After computing response rate and response quality measures for all studies, we retained 3815 respondents for whom all information was available. The other respondents were of screened-out since the project time was expired or because they were not suitable for the specific questionnaire. The respondents were divided into 42.1% male and 57.9% female respondents. Most of the respondents were between 26 and 55 years old (19-25: 8.9 %; 26-35: 21.2 %; 36-45: 24.0 %; 46-55: 35.5 %; 56-65: 12.8 %; <65: 1.7%). In terms of employment, 11.1% of the respondents worked for the government or other public authorities, 9.5% worked in healthcare, 7.4% in education, 7.1% were not professionally active, 5.3% were students, and all others were equally divided over different industries. The majority of respondents was married with children (45.6%), followed by married respondents without children (19.3%), single without children, (12.2%), and single with children (8.5%). 10.5% lived with grandparents/family.

Results

For multivariate kurtosis both Mardia's (1970) and Srivastava's (1984) test indicated a significant deviation from multivariate normality (for both tests $p < 0.001$). Similarly, Srivastava's test of multivariate skewness showed a significant deviation from multivariate normality ($p < 0.001$). As a consequence of the deviation from multivariate normality, we used the Satorra-Bentler-scaled χ^2 statistic (χ^2_{SB}) with robust standard errors for the estimation (Satorra and Bentler 1994). Confirmatory factor analysis of the SPI, effort and response quality, provided an adequate model fit ($\chi^2_{SB} (695) = 14606.256$; RMSEA=0.072; NFI=0.950; NNFI = 0.946; CFI=0.952; IFI = 0.952; RFI = 0.944). By examining the residuals, the modification indices and loadings, eight problematic items, one for every construct, could be identified and excluded. The final confirmatory factor analysis provided a good model fit ($\chi^2_{SB} (360) = 6205.834$; RMSEA = 0.065; NFI = 0.964; NNFI = 0.959; CFI = 0.966; IFI = 0.966; RFI = 0.957). Composite reliabilities ranged from 0.654 to 0.895 (interest 0.868; enjoyment 0.890; curiosity 0.811; give opinion 0.852; helping 0.895; recognition 0.654; incentives 0.802; obligation 0.690; effort 0.808; response quality 0.755).

Next, we assessed discriminant validity between all eight constructs of the SPI as well as self-reported effort and response quality. We constrained the estimated correlation parameter (Φ_{ij}) between each pair to one and then performed a Satorra-Bentler χ^2_{SB} difference test (Satorra and Bentler 2001) on the values obtained for the constrained and unconstrained models (Anderson and Gerbing 1988). Assessing the discriminant validity of the SPI with effort and response quality showed that the unconstrained coefficients were superior to the constrained coefficients (effort: $\Delta \chi^2_{SB} (1) = 67.67$, $p < 0.001$; response quality: $\Delta \chi^2_{SB} (1) = 103.35$, $p < 0.001$). In addition, discriminant validity between all factors in the SPI was examined. Again, the unconstrained coefficients were superior to the constrained coefficients ($p < 0.05$ for all Chi-square difference tests). We also examined discriminant validity of the SPI as recommended by Bagozzi and Warshaw (1990), who state that discriminant validity between two factors is achieved when their correlation is less than 1.0 by an amount greater than twice its standard error. This criterion was met indicating discriminant

validity between all items in the SPI. Therefore, we conclude that the SPI exhibits discriminant validity.

To test the impact of the different response motives on response rate and response quality, canonical correlation was performed between the SPI and the set of response quality variables using the CANCELL procedure in SAS release 9. We employed canonical correlation analysis since it is the most appropriate and powerful multivariate technique in situations with multiple dependent and independent variables. Canonical correlations present the general model on which many other multivariate techniques are based.

With all eight canonical correlations included, $F(64, 14599) = 13.39, p < 0.001$. With the first four canonical correlations removed, F -values were not significant anymore ($F(16, 7742.2) = 1.38, p = 0.140$). Therefore, the first four pairs of canonical variates accounted for the significant relationships between the SPI and response quality. The first canonical correlation was 0.42 (18% overlapping variance) and the second was 0.28 (8% overlapping variance). The third canonical correlation was 0.18 (3% overlapping variance) and the fourth was 0.12 (1% overlapping variance). Although significant, neither of these last two canonical correlations represented a clearly interpretable relationship. Since interpretation of the third and fourth canonical correlations and their corresponding variates were marginal, the results are not reported here.

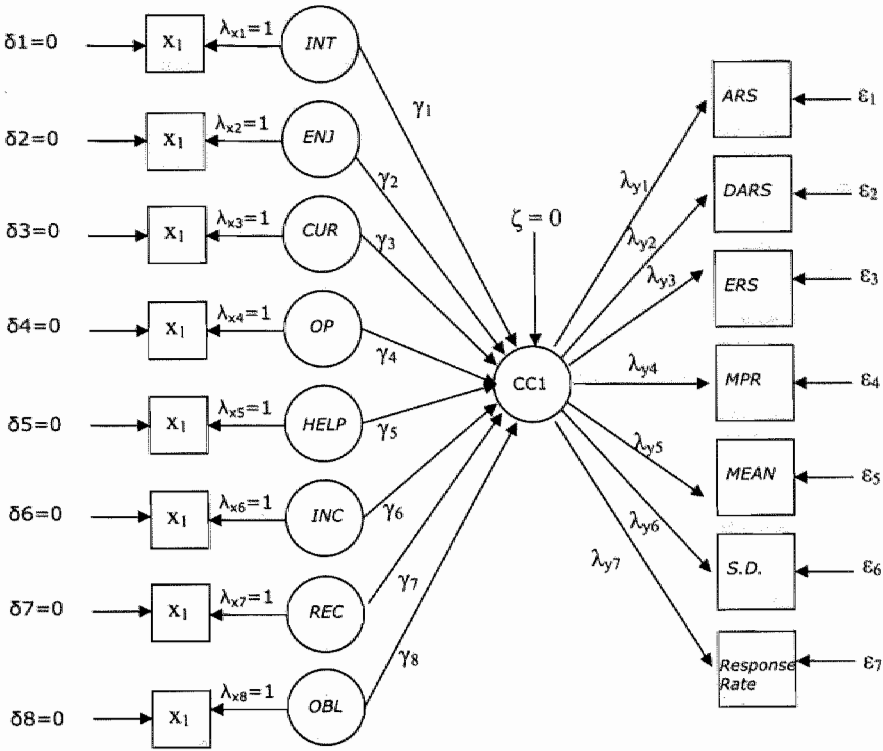
Table 5.3 shows the correlations between the variables and the canonical variates, standardized canonical variate coefficients, within-set variance accounted for by the canonical variates (percent of variance), redundancies, and canonical correlations.

One key issue in the interpretation of canonical correlation analysis concerns weight instability (Lambert and Durand 1975). To evaluate weight instability, the total sample was split into an analysis and a validation sample. Canonical correlation was then conducted with each subsample, thus deriving separate weights for each group. Then, estimated raw canonical coefficients of the analysis sample were multiplied with the predictor and criterion set of the validation sample to derive additional variate estimates. These estimates were correlated with the initial variates for the validation sample. The correlations for the first and second variate for the predictor (0.891; 0.856) and criterion set (0.867; 0.696) were substantial in magnitude and statistically significant ($p < 0.001$).

In order to formally test the significance of the canonical correlation coefficients, we employed structural equation modeling using an approach suggested by Bagozzi, Fornell and Larcker (1981) (see Figure 5.2). We were able to replicate the weights of the findings and our results suggested that for the first variate, INT, ENJ, OP, HELP, INC, ARS, DARS, ERS, MPR, ISM and ISSD were significant ($\chi^2_{SB} (42) = 316.898$; RMSEA=0.0515; NFI=0.986; NNFI = 0.970; CFI=0.988 IFI = 0.988; RFI = 0.966). However, canonical coefficients are subject to instability and should be interpreted with caution. To overcome the deficiencies inherent in canonical coefficients (Thompson 1984), canonical correlations have been increasingly used as a basis for interpreting the results. The dominant correlations of the first and second variate reveal that the first variate consists of intrinsic motivation while the second is composed of extrinsic motivation (see Table 5.3). The self - other distinction does not seem to discriminate between the two variates. The results for the first variate indicated that respondents who are intrinsically motivated - interest (0.962), enjoyment (0.665), curiosity (0.759), giving an opinion (0.715), and helping (0.710) - are associated with more yea-saying (0.840) and higher means (0.813).

Concerning the second variate, CUR, OP, REC, DARS, and S.D. were statistically significant at $p < 0.05$ ($\chi^2_{SB} (34) = 124.118$; RMSEA=0.0281; NFI=0.995; NNFI = 0.991; CFI=0.996; IFI = 0.996; RFI = 0.987). A look at the correlations of the second variate showed that the second variate was composed of all extrinsic motives (incentives, need for recognition, and obligation). The corresponding canonical variate from the response quality set consisted of DARS, ERS, and ISSD. Taken as a pair, these variables suggested that incentives (-0.446), need for recognition (-0.618), and obligation (-0.309) are associated with less nay-saying (0.807), less extreme responding (0.663), and lower standard deviations (0.839). Midpoint responding (-0.412) and means (-0.368) were somewhat higher.

Figure 5.2: Canonical Correlation as a MIMIC Model – First Variate Pair^a



^aCovariances between error variances (ϵ_1 - ϵ_7) of the reflective items for the canonical variates are not depicted although they are specified in the model

Table 5.3: Correlations, Standardized Canonical Coefficients, Canonical Correlations, Percents of Variance and Redundancies between the SPI and Response Rate and Response Quality

	First Canonical Variate		Second Canonical Variate	
	Correlation	Coefficient	Correlation	Coefficient
SPI Set				
INT	0.962	0.747**	-0.137	0.016
ENJ	0.665	0.107*	0.033	0.158
CUR	0.759	0.019	-0.328	-0.518*
OP	0.715	0.174*	0.306	0.818*
HELP	0.710	0.198**	0.003	0.153
INC	0.186	0.090*	-0.446	-0.169
REC	0.416	-0.106	-0.618	-0.723*
OBL	0.503	-0.081	-0.309	-0.176
Percent of Variance	0.427		0.112	Total = 0.540
Redundancy	0.075		0.009	Total = 0.084
Response Set				
Response rate	0.163	0.222**	0.125	-0.096
Response Quality				
ARS	0.840	0.465**	0.147	-0.018
DARS	-0.327	-0.376**	0.807	0.500*
ERS	0.241	0.232**	0.663	-0.004
MPR	-0.218	-0.110**	-0.412	-0.167
MEAN	0.813	0.385**	-0.368	-0.075
S.D.	0.325	0.179**	0.839	0.584*
RQ	0.222	0.000	0.506	0.000
Percent of Variance	0.198		0.268	Total = 0.466
Redundancy	0.035		0.021	Total = 0.056
Canonical Correlation	0.419		0.2777	

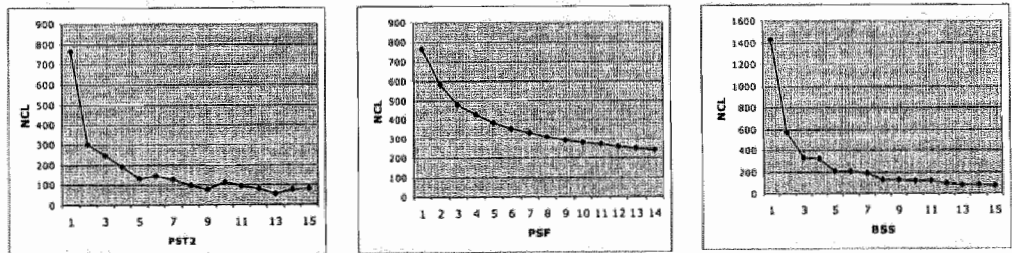
**significant at $p < 0.01$ based on SEM*significant at $p < 0.05$ based on SEM

In addition to examining the impact of the typology on response rate and response quality, a cluster analysis was carried out on the basis of the different response motives as identified in the typology. Initially, the raw SPI scores were used as input for clustering. Yet, this resulted in a cluster solution where the level of motivation (low, medium or high levels of motivation) dominated over the response motives. This indicated that the response style effects were strongly present, also in the assessment of the SPI scale. In this situation, Schaninger and Buss (1986) recommend that response-style effects are removed by standardizing the data to get a clearer and more meaningful cluster solution. Standardization is appropriate when (1) the interpretability improves, (2) the ability to clearly contrast clusters improves, (3) clear-cut differences in relative ratings are demonstrated within cluster profiles, (4) a priori segments are

confirmed, and (5) statistically stronger relationships of cluster membership to key outcome variables are found. Since all five criteria were met, we standardized our data. We employed within-individual standardization as method of adjustment since it corrects for between-individual differences in variance and range as well as central tendency and is preferred over row centering and normalization (Schaninger and Buss 1986).

Clustering was conducted in a two-step approach (Punj and Stewart 1983). First, the CLUSTER procedure in SAS Release 9 was used to conduct a hierarchical cluster analysis using Ward's minimum variance method with squared Euclidian distances. There are no completely satisfactory stopping rules in determining the optimal number of clusters. Yet, Milligan and Cooper (1985) verified that the pseudo F-statistic by Calinski and Harabasz (1974) and the pseudo t^2 -statistic related to the $J_e(2)/J_e(1)$ statistic by Duda and Hart (1973) perform well. Therefore, the pseudo F-statistic, the pseudo t^2 -statistic and the fusion coefficients (between sum of squares for Ward's minimum variance method) were plotted against the number of clusters. Inspecting these plots suggested a three cluster solution.

Figure 5.3: Pseudo t^2 -statistic, F-statistic, Fusion Coefficient



Second, the input from the hierarchical cluster analysis was used as input for the nonhierarchical cluster analysis. In accordance with the procedure recommended by Punj and Stewart (1983), the 3815 cases were randomly split into an analysis sample and a validation sample. The analysis sample generated possible alternative cluster solutions and used the centroids of the hierarchical cluster analysis as starting seeds for the K-means clustering algorithm. The FASTCLUS procedure in SAS Release 9 was utilized, which employs nearest centroids sorting.

Next, the validation sample was used for cross-validating the different cluster solutions. First, each observation was assigned to the nearest seed and the cluster seeds were updated. Second, each observation was only assigned and seeds were not updated. Thereafter, the two different cluster solutions were compared. Milligan and Cooper (1986) recommended the adjusted Rand index as the index of choice. As can be seen in Table 5.4, the adjusted Rand index for the three and four cluster solutions are very high. Since the pseudo F-statistic, the pseudo t^2 -statistic and the fusion coefficients indicated a three cluster solution and the results proved to be richer in contrast and content, we chose a three cluster solution.

Table 5.4: Adjusted Rand Index

# of Clusters	2	3	4	5	6
Adjusted Rand Index	0.8510	0.923	0.981	0.724	0.580

To verify the three cluster solution and to examine whether the SPI reliably separates the different clusters, we performed discriminant analysis. Examination of the normality of the data revealed no serious violations of univariate normality. In a first step, we examined whether discriminant analyses replicated and validated the assignment of individuals to the respective clusters based on the SPI scores. Thus, cluster membership served as a dependent variable for the discriminant analysis. A comparison of the discriminant and cluster solutions revealed 90% agreement. In addition, we examined the predictive validity of the discriminant analysis in grouping individuals based on the SPI scores. This cross-validation, where we used an analysis and a validation sample, resulted in an agreement score of 99%. Hence, the three clusters significantly differed along the SPI dimensions and SPI scores could be used to reliably assign individuals to the proper cluster.

Now that the number of clusters has been determined, the data were pooled and used as input into K-means cluster analysis. Moreover, the clusters were subjected to a series of ANOVA's using the Games-Howell post-hoc test to compare the different groups as implemented in SPSS release 13. In addition to self-assessed effort and response quality, ISM and ISSD, the four response styles (ARS, DARS, ERS, MPR) were added up to derive at an overall assessment of response style. Response likelihood as well as most aspects of response quality

were significantly different at $p < 0.05$. A description of the final clusters and the ANOVA results are displayed in Table 5.5.

Based on the differences that emerged from these analyses, respondents in the three clusters can be described as follows:

1. **Voicing assistants:** 27% of the sample is mainly motivated by intrinsic – other response motives. Both giving an opinion and helping are evaluated as being extremely important and are the only dominant response motives for this cluster. All other response motives are around zero or negatively related. Response rates are lower than for the third cluster (81.86%), but respondents have a neutral response style.
2. **Reward Seekers:** Respondents in the second cluster (25%) mainly participate because of the incentives that they receive in turn for their participation. All other motives except giving an opinion (.366) are negatively related. This cluster has the lowest response rate (80.22%) and lowest levels of self-reported effort and performance. Respondents tend to engage in nay-saying, even though this difference was not statistically significant. Overall, response style is best for this cluster. Respondents in this cluster are youngest (38.58 years), only 56.7% have children, and the cluster contains more students.
3. **Intrinsics:** Respondents in the largest cluster (48%) are motivated by multiple intrinsic motives. While enjoyment and giving an opinion are the strongest response motives, also interest, curiosity and helping are positively related. Incentives are highly negatively related and are therefore not important at all. This cluster has the highest response rate, but also the worst response style. Respondents in this cluster are yea-sayers, have higher means and standard deviations, and therefore the worst overall response style. Respondents in this cluster are older, have more children, and more respondents are not professionally active.

5.4 Discussion

This study aims to extend the sparse literature on online access panels by examining what motivates panel members to participate in online surveys. In addition, panel effectiveness parameters were examined by linking response motives to response rate and response quality and by clustering respondents. In the following paragraph, we conclude the discussion of the empirical findings by relating them to the three objectives of this paper.

TABLE 5.5: Cluster Descriptors and Results

	Cluster 1 Voicing Assistants		Cluster 2 Reward Seekers		Cluster 3 Intrinsics	
	1015		945		1784	
Means:	Stand.	Unstand.	Stand.	Unstand.	Stand.	Unstand.
Interest:	-.311	4.000	-.210	3.800	.245	5.140
Enjoyment	.043	4.522	-.437	4.735	.500	5.554
Curiosity:	-.289	3.982	-.032	4.050	.323	5.265
Give opinion:	.635	5.374	.366	4.644	.548	5.634
Helping	.652	5.388	-.086	3.971	.335	5.316
Incentives:	-.230	4.077	.609	4.890	-.818	3.475
Need for recognition	-.592	3.574	-.583	3.280	-.823	3.479
Obligation:	-.094	4.550	-.500	3.360	-.309	4.260
Age:	40.66 ^A		38.58 ^B		41.32 ^C	
Response rate	81.86% ^A		80.22% ^A		85.26% ^B	
Response Quality						
Effort	4.52 ^A		4.38 ^B		4.52 ^A	
Performance	4.86 ^A		4.65 ^B		4.91 ^A	
- Accuracy						
ARS	-.015 ^A		-.036 ^A		.280 ^B	
DARS	.095 ^A		.134 ^A		.042 ^A	
ERS	.036 ^{A,B}		-.025 ^A		.129 ^B	
MPR	.068 ^A		.048 ^A		-.026 ^A	
RS	.183 ^{A,B}		.120 ^A		.426 ^B	
MEAN	-.102 ^A		-.112 ^A		.202 ^B	
S.D.	-.017 ^A		.019 ^A		.116 ^B	

^{A,B,C} different subscripts indicate that the clusters are significantly different from each other at $p < 0.0$

First, the survey participation inventory was generated to measure response motives as identified in the typology. The SPI was thoroughly developed and presents a cohesive tool that reliably measures panel members' motivation to participate in online surveys. Second, the SPI was linked to important response parameters in order to examine whether they have a differential impact on response rate and response quality. The canonical correlation results confirm a relationship between intrinsic motives, yea-saying and higher means. Respondents who are intrinsically motivated seem to be so eager to participate that they are more favorable in their answers.

Table 5.6: Cluster Demographics

	Cluster 1 Voicing Assistants	Cluster 2 Reward Seekers	Cluster 3 Intrinsics
	1015	945	1784
Age:	40.66 ^A	38.58 ^B	41.32 ^C
Gender			
- female	56.3%	57.8%	58.9%
- male	43.7%	42.2%	41.1%
Children			
- yes	61.9%	56.7%	67.90%
Top Four Occupations			
#1	Public sector employed (11.0%)	Healthcare and social services (10.2%)	Public sector employed (11.9%)
#2	Education (9.0%)	Public sector employed (9.8%)	Healthcare and social services (9.8%)
#3	Healthcare and social services (8.9%)	Education (9.2%)	Other (9.8%)
#4	Other (7.3%)	Other (7.7%) and students (7.7%)	Not professionally active (7.9%)

^{A,B,C} different subscripts indicate that the clusters are significantly different from each other at $p < 0.0$

Apparently, intrinsic motives involve pleasurable internal feelings or thoughts that are carried forward to the evaluation of products, concepts, ideas, or attitude that are inquired in market research. Extrinsic motives on the other hand are associated with less nay-saying and less extreme responding. In addition, these respondents exhibit lower standard deviations and more midpoint responding. Extrinsically motivated respondents thus seem to be attracted by the valence of the reward and therefore engage in satisficing by using a more neutral response strategy with more midpoint responding.

Third, we wanted to investigate the possible existence of respondent clusters that are driven by different underlying motives to participate. Indeed, three clusters could be clearly identified by using the SPI. The cluster analysis revealed that intrinsics are more likely to participate than reward seekers or voicing assistants. However, intrinsics also had the worst response style. Respondents in this cluster are yea-sayers, more extreme in their answers, and show higher means and standard deviations. A possible explanation could be the demographic profile of this cluster. Existing evidence verifies that women are more intrinsically motivated and men extrinsically motivated (Holbrook 1986). Also respondents in cluster three are predominantly female, relatively older, have more children, and are not professionally active. They seem to be so averse to

participate that they are more positive in their answers. In any case, the evidence indicates that researchers indeed need to trade-off high response rates and lower response quality.

Interestingly, the incentives cluster is the smallest cluster which implies that incentives do not seem to be such an important response motive after all. These findings are in contrast to the existing survey literature, which suggests that incentives are the most important driver of response. Respondents who are mainly motivated by incentives engage somewhat more in midpoint responding, and have the lowest level of self-reported effort and response quality. It seems that people that respond for the love of money or other rewards love earning it the easy way. However, overall the response quality is actually relatively good and comparable to cluster one.

5.5 Theoretical and Managerial Implications

This paper contributes to the understanding of why people participate in online surveys in several important ways. As mentioned earlier, most researchers have focused on the design of surveys and have neglected individual differences in the type of motivation that stimulates respondents to participate. Using the newly developed SPI, we developed a more complete typology which allows the investigation of more complex relationships between types of motives and response behavior. For example, this study provides more insights into how incentives are actually working by untangling the complex interrelation of multiple response motives. Moreover, response rate and response quality were tested using both self-reported measures as well as objective variables such as participation behavior in the past year or response style and quality. This strengthens the findings of this study and helps to overcome common method bias.

The survey participation inventory presents a very interesting tool that market researchers can use to measure the type and strength of motivation of their respondents. The type of motivation can then be matched with the corresponding design stimuli, for example by using statement such as "your opinion counts" in the subject line of e-mail invitations, wording of invitations, or the design of the surveys. If the SPI is not administered a-priori, researchers should at least change the nature of appeals over successive waves (Groves, Singer, and Corning 2000) in order to reach a wider audience with individual

differences in the dominant response motives. In addition, researchers can also use the SPI for the recruitment of new panel members. Respondents can be screened on their type of motivation in order to invite the 'right' respondents to become a panel member. However, the empirical evidence shows that the definition of 'right' is actually much more complex than initially assumed. The results show researchers have to trade-off high response rates, which are supposed to reduce nonresponse error, a potential loss in response quality. Also respondents who are mainly driven by incentives seem to be somewhat problematic since they seem to engage in some form of satisficing.

The representativeness of online surveys is a reoccurring question to everyone conducting online surveys. Yet, till this day, representativeness has only been investigated in terms of demographics where sophisticated weighting schemes are used nowadays to match sample and population of interest. The findings show that even though an online sample might be representative in terms of demographics, respondents may be overly motivated or strongly driven by particular response motives, which can have a severe impact on results. Consider the case of highly intrinsically motivated respondents which show acquiescence, more extreme responses, higher means and standard deviations. Knowing that the respondent was highly intrinsically motivated would already be beneficial to interpret the results cautiously and put them into perspective.

5.6 Limitations and Future Research Suggestions

One limitation of this paper is that we use a self-reported measure to assess the importance of the different types of motives, which could be influenced by a social desirability bias. We tried to rule this out by examining the Marlowe-Crowne social desirability scale in the second pretest (Crowne and Marlowe 1960) where no significant impact on the SPI was found. An experiment would have overcome this problem. Yet, we purposefully chose this design since we were specifically interested in why and how response stimuli work and which underlying motivations are triggered and therefore wanted to go beyond the experimental designs in the traditional survey literature. Similar studies such as the measurement of the work preference inventory by Amabile et al. (1994) were also conducted via survey which was not considered problematic.

It would have been interesting to examine the interaction effect of past participation, survey topic, and lifestyle variables. This would have provided us

with an even better understanding of response behavior. Unfortunately, the database of the panel with which we cooperated was just created recently and they were not able to provide us with this information. With the advancement of panel management tools, this information will be available in the future and thus presents an interesting avenue for further research.

The focus of this study was on online surveys conducted through online access panels. It would be extremely interesting to examine in how far the results transfer to other survey modes. While we think that the SPI could also be applied to offline modes, we expect differences in the importance of several motives. It would also be interesting to compare the type and level of motivation between online, mail and telephone surveys, especially if considering the current drastic decreases in response rates for mail and telephone surveys (Morin 2004). We suspect that respondents in online panels are probably more motivated to participate than respondents to telephone, mail, or mall intercepts since they actively decided to join an online panel. Lower levels of motivation would probably decrease response quality since respondents are more likely to satisfice (Krosnick 1991). Future research could therefore examine the type and level of motivation in traditional survey modes and verify the trade-off between the type of motivation, response rate and response quality.

This paper focuses on response motives. Similar to Herzberg's hygiene and motivation factors, we believe that a different set of factors is responsible for non-response. An exploratory question in the questionnaire shows that the factors driving non-response are very diverse and that more situational variables play a role. The factors that were mentioned as a reason why someone did not respond were: (1) too busy (50.8%), did not belong to target group (40.5%), subject was uninteresting (33.3%), duration of survey had expired (29.2%) I knew nothing about the subject (21.8%), questionnaire failed to start (20.7%), and I was on a holiday (17.4%). Future research should expand these preliminary findings and investigate why people do not participate in (online) surveys.

Finally, the empirical studies were only conducted in one country. The findings of Albaum, Evangelista and Median (1998) as well as Evangelista, Albaum and Poon (1999) suggest that there are differences between Europe, Asia-Pacific, and Northern America. Future research should therefore extend this study by examining more countries with different cultures.

CHAPTER 6: CONCLUSION

The final chapter of this dissertation consists of a discussion of the results of the four different studies, their managerial and theoretical implications, a conclusion, and a perspective on future research.

"The two most significant advances in survey methodology during the twentieth century were the introduction of random sampling in the 1940s and interviewing by telephone in the 1970s. Both of these innovations have transformed how most major surveys are done. We are now witnessing another development in survey methodology, the consequences of which may prove to be even more profound. It is the collection of survey data through [...] e-mail [and] the World Wide Web."
(Dillman 2000, p. 352)

6.1 Synopsis

As the quote illustrates, Dillman (2000) views the development of online surveys as one of the most profound developments in the survey research industry. This potential of online surveys is driven by two main developments: (1) The rapid development of the World Wide Web where over half a billion people worldwide now have Internet access (NUA Internet Surveys 2003) and (2) the advantages of online surveys, especially low costs and fast responses, which make online surveys an effective alternative to traditional survey methods. Metzke and Allan (2005) even predict that within two years time, online surveys will have an average predicted share of business of 39.6%, more than telephone (30.7%), face-to-face (23.9%) or mail surveys (5.7%). Despite the increasing popularity, scientific research on online surveys is still scarce. This is surprising since online surveys are more than just a new survey mode. According to Miller and Dickson (2001), online surveys present a technological and cultural change that influences what we do as researchers and how we think about research.

As the success of online surveys will "stand or fall by the credibility of the data it generates" (Taylor 2000, p.51) in relation to data collection costs, empirical evidence is needed that examines critical factors and their impact on the effectiveness of online surveys. Therefore, the overall objective of this dissertation was to investigate the impact of design factors, mode effects, and individual differences on the effectiveness of online surveys. The studies were conducted in a B-2-B as well as B-2-C context using surveys, experiments, and Generalizability theory (G-theory). The types of web surveys that were investigated were volunteer opt-in panels and mixed mode surveys conducted with list-based samples. The different perspectives on online surveys were combined in the following overall research question:

What are the critical factors that influence the effectiveness of online marketing research?

This main research problem was analyzed by conducting four consecutive studies, which were discussed in chapter two through five. In chapter 2, we examined the effect of the timing of follow-ups, different incentives, length, and presentation of the questionnaire on the response rate and response quality in an online experimental setting. Systematic differences in response quality were investigated by examining and comparing the equivalence (chapter 3) and generalizability (chapter 4) of online and mail surveys. Chapter 5 focused on individual differences in motives to join an online access panel and the corresponding effects of response motives on response rate and response quality.

The final chapter of this dissertation consists of a discussion of the results of the four different studies and a perspective on future research.

6.2 Discussion

This dissertation contributes to the survey literature in several important ways. As illustrated throughout this dissertation, the virtual environment of online surveys has changed the surveying process significantly. Because of these different processes, empirical findings from the existing survey literature cannot per se be generalized to the online setting. Also, existing studies on online surveys are still scarce. To fill the lacunae regarding quality of online surveys, we conducted several interrelated studies that present a comprehensive assessment of critical factors that influence response quality of online surveys using different empirical methodologies. The four studies in this dissertation build upon each other and answer successive research questions which, taken together, help researchers in three different phases of the surveying process: (1) creating an optimal design of online surveys that increases response rates and response quality, (2) comparing the quality of online and mail surveys, and (3) evaluating the quality of online access panels.

6.2.1 Design of Online Surveys

Study one examined the impact of several important design factors where the evidence from the traditional survey literature could not be translated to the online environment. The virtual setting requires, for example, different incentive schemes since cash incentives cannot be attached to an online questionnaire. It is also unclear how a 'long' online survey is defined. Rosenblum (2001), for example, states that online surveys should not contain more than 20 questions, which is generally considered too short for most surveys. To answer the fundamental question of how to design online surveys, we conducted an online experiment in which the effect of the timing of follow-ups, different incentives, length, and presentation of the questionnaire on the response rate and response quality was examined.

The results showed that long questionnaires still generated a surprisingly high response, which indicates that online surveys have the potential to be used for substantive research. Moreover, we found that lotteries with several small prizes are very effective in increasing the response rate, especially in short surveys. This is surprising since the mail survey literature finds that small cash incentives for every respondent are more successful in increasing response rates than lotteries or donations (Warriner et al. 1996).

In addition, study one was also one of the first studies to examine the use of design elements that make use of the interactive nature of the Internet, such as product images. Showing pictures of products instead of only giving the name increased response quality while response rates were lower, especially in the long version. Also, respondents in the visual version preferred an incentive that compensated for the time and effort they spent on the questionnaire rather than a donation to charity. This indicated that respondents in the visual version were subject to longer downloading times.

6.2.2 Comparability of Online and Mail Surveys

Once researchers know how to design online surveys, it is paramount to investigate the response quality of online surveys and examine whether there are any systematic differences between online and offline self-administered surveys. Because of the importance of response quality, two studies in this dissertation examined different aspects of response quality. In both studies, we conducted a large-scale real life quasi-experiment in which relevant influencing

factors such as sample source were the same across groups so that systematic differences between online and mail surveys could be examined.

Study two examined whether online and mail surveys produce convergent results. Even though online surveys seem to replace telephone surveys in the commercial world, online and mail surveys are similar in structure and are therefore often used in mixed-mode service quality studies. In the context of a large B2B service quality assessment, we analyzed the accuracy and completeness of respondents' answers to both open and closed questions. Overall, the results suggested that online and mail surveys produce equivalent results. Both the composite reliability and the average variance extracted showed consistently high levels for both groups, and the means and variance-covariance matrices were equal across modes. Convergent validity between the two survey modes was supported further by a measurement-invariance test based on multigroup confirmatory factor analysis (Jöreskog 1971). However, minor differences occurred in the answers to open-ended questions. For example, online respondents provided more improvement suggestions, indicated more often to which competitor they wanted to switch, and provided lengthier answers in response to requests for examples of positive experiences with the company.

Study three compared the generalizability of online and mail surveys in a cross-national service quality study. We illustrated how G-theory can be used to compare online and mail surveys while taking data collection costs into account. Looking at the required or desired response quality in relation to data collection costs is interesting since the two are directly linked. For example, if more items are added to a questionnaire to improve the reliability of the constructs and thus the response quality, the cost must also increase. Even though online surveys benefit from lower costs, the costs of large-scale international performance measurement, even when administered online, may still be substantial. Surveyors can thus pursue two different goals: (1) maximize response quality given a certain cost level or (2) minimize costs given the desired response quality level. The results of this study indicated a comparable level of generalizability between online and mail surveys, confirming the findings of study two. The cost functions do not only prove that online surveys benefit from considerably lower costs, but help researchers to decide how restricted research resources may best be employed.

Both studies find convergent results between online and mail surveys. These findings are promising and direct researchers' attention to other influences on the quality of online surveys, such as the sample. Study two and three also bear strong implications for mixed-mode surveys since they show how online and mail surveys can be compared.

6.2.3 Quality of Online Access Panels

This dissertation shed more light on the design of online surveys and provided evidence that online surveys are comparable to mail surveys. While this is important information that helps in the set-up of online surveys, the quality of online studies stands and falls with the quality of the sample. In a B-2-B context, many companies still have a database from which they can draw a random sample. In contrast, most B-2-C companies have to rely on e-mail addresses from online panels. Though a majority of online panels are very large and apply sophisticated weighting techniques to compose a representative sampling frame with respect to demographics, empirical evidence suggests that on- and offline samples are still different, leading to incomparable results (e.g., Robinson, Neustadt, and Kestnbaum 2002; Sparrow and Curtice 2004; Vehovar, Lozar Manfreda, and Batagelj 1999). These differences could be caused by self-selection of online panel members.

Since panel members have to be highly motivated to join an online panel and to continuously participate in online surveys, online access panels are likely to contain respondents which are more willing to cooperate. Many respondents are members of several panels or participate because of the money they can win (e.g. "The free Get Paid to Take Surveys Online Guide;" <http://www.surveys4money.com/>). As mentioned in the introduction, ComScore Networks (2005) even report that 30% of all online surveys are completed by only 0.25% of the Internet population.

Therefore, the aim of study four was to enhance the existing survey literature by identifying the underlying motives that stimulate respondents to participate in online panels. This dissertation presents one of the first studies to examine individual difference in response motives. Specifically, we established a more complete typology of response motives which was tested through the newly developed SPI. We also examined the effect of the type of motivation on the participation rate as well as response quality. Based on the SPI, respondents

could be divided into three clusters where respondents can be characterized as voicing assistants, reward seekers, and intrinsic. Intrinsic exhibited higher participation rates but lower response quality. This suggests that researchers have to trade-off increased response rates which are supposed to reduce nonresponse error with a potential loss in response quality due to an overrepresentation of highly motivated respondents.

The SPI presents a useful tool for researchers to assess respondents' motives to join online access panels or participate in specific online surveys. Researchers can thus determine whether their sample is representative with respect to the type and level of motivation and whether they have overly motivated or professional respondents in their sample/panel. In addition, the design of the e-mail invitation and survey can be matched to the response motive in order to increase response rates when inviting panel members to participate in surveys.

6.2.4 Results in Reference to the Overall Effectiveness of Online Surveys

The results of this dissertation can be referred back to the effectiveness equation as presented in the introduction of this dissertation. In essence, effectiveness of marketing research can be expressed as a function of two variables: (1) costs and (2) data quality. One of the key advantages of online surveys is lower costs (e.g., Illieva, Baron, and Healey 2002). Yet, claims regarding the cost-effectiveness of online surveys in international service research were largely based on anecdotal evidence and had not been systematically analyzed relative to mail surveys. To compare the cost structures of online and mail surveys, we conducted in-depth interviews with five researchers working at three different international market research agencies. We came up with two cost functions for mail and online surveys (details on how the cost functions were derived can be found in Table 4.2). Our equations show that the fixed costs of online surveys are slightly higher because of the programming efforts upfront, but variable costs are virtually zero. Thus, online surveys are indeed less expensive than mail surveys

Costs do not only include money but also time spent during the complete research cycle. As mentioned earlier, online surveys are reported to have much faster response times than for example mail surveys. In study one, we received

more than half of the final responses (52.9%) after only three days and the average response time was 6.6 days. Even though slightly higher than the 5.59 days as reported by Illieva, Baron and Healey (2002), our evidence confirms the speed of online surveys. The response time was not recorded in the other studies, but our experience verifies the fast response time of online surveys. One interesting finding from study one is that respondents in the lottery group responded 1 and 1.5 days faster than the donation and voucher group, respectively. This could be due to the fact that respondents in the lottery might have inferred that by responding early, they have a higher chance of winning a price.

Finally, all online surveys should be viewed in light of potential coverage and sampling, nonresponse and measurement error. As outlined by Couper (Couper 2000), different types of web surveys are more or less sensitive to the different types of error. Companies should be careful when using volunteer opt-in panels. While these panels are growing fast and receive widespread attention, study four shows that there is a self-selection of volunteers who join the panel.

Since this dissertation provides new empirical evidence on costs and quality of online surveys, researchers are now able to assess the effectiveness of online surveys and chose the most appropriate research mode for their research objective at hand.

6.3 Perspectives on Future Research

This paragraph provides general directions for future studies in marketing research and the progression of online surveys. As mentioned in the introduction, there has been growing skepticism about the role and value of traditional marketing research in influencing managerial decision making. Globalization, technology and communication, reduction in traditional boundaries, and a strategic focus on the use of information seem to require a new era of marketing research. Common points of criticism about traditional survey modes are high costs and long research cycles in relation to the quality of the information that is delivered (Burke, Rangaswamy, and Gupta 1999). The advances of the Internet offer an opportunity to address these limitations of existing research approaches. Online surveys are cheaper, faster and easier to complete, and foster large scale, international research.

Furthermore, the Internet democratizes the survey-taking process since large scale data collection is no longer restricted to major organizations. The relatively low cost of conducting online surveys essentially puts the tool in the hands of almost every person with access to the Internet (Couper 2000). Yet, the downside is that the design and development of online surveys has in large part come from computer programmers, many of whom have little or no training in survey methodology (Dillman and Bowker 2001). The flood of online survey tools makes it increasingly difficult to distinguish the good from the bad. Survey methodologists are in demand to bring their knowledge on methodology to online surveys.

Study two and three of this dissertation compared online and mail surveys because the type of research for which mail surveys are used is most likely to be supplemented by online surveys. However, future research should examine how alternative modes compare to online surveys. It might be especially interesting to compare telephone surveys, which have the largest share of business at the moment (34.4%) (Metzke and Allan 2005). Telephone surveys are not self-administered and random digit dialing makes it possible to obtain a representative national sample. However, telephone surveys are threatened by growing time pressure, rise of telemarketing, more mobile phones and less fixed lines, answering machines, caller ID, call blocking and no answers (Morin 2004).

Future studies should also examine additional design factors, such as survey invitations and introductions or different layouts. Also the impact of potential moderating factors on the results should be investigated. In addition, systematic differences between online surveys and traditional modes should be verified in a B-2-C context where sampling is a much bigger issue.

Nonrespondents have received little attention in the survey literature. Brennan and Hoek (1992) find that refusers differ from nonrespondents with regard to their opinion toward survey participation and demonstrate quite different attitudes and behaviors. Also Bickart and Schmittlein (1999) propose that some respondents are less prone to participate in surveys. This problem might be intensified for online surveys where the threshold for participation is higher; respondents need to be online and be willing to devote their precious time to participate in online surveys while they could spend their online minutes on more enjoyable things. Since nonresponse is likely to be driven by different factors than the response motives as identified in study four, future research

could identify reasons that drive nonresponse and the impact on sample representativeness and response quality. Researchers could examine in how far nonrespondents affect coverage, sampling, and nonresponse error.

This dissertation used volunteer opt-in panels and conducted mixed mode surveys with list-based samples. Future research could also examine other types of online surveys (Table 1.1). In addition, online qualitative research might be worth investigating. Several authors and practitioners acknowledge the wide range of possibilities for qualitative online research (e.g., Montoya-Weiss, Massey, and Clapper 1998; Scholl, Mulders, and Drent 2002). However, the role of the moderator differs and spontaneous reactions and non-verbal cues are difficult to simulate in an online environment (e.g., Greenbaum 1998). As Greenbaum (1998) articulates that "Internet Focus Groups are not Focus Groups - So Don't Call Them That", future research should assess whether online qualitative research methods still provide detailed insights into the phenomena under study. Hence, researchers should critically compare several online and traditional qualitative research methods to assess their quality.

While the interactivity of the Internet offers great design opportunities, most online surveys are mere copies of their paper counterparts. Market research companies are only slowly starting to implement more interactive question formats, for example by using alternative interfaces or metaphors. Advancing the design of online surveys could be a key competitive advantage that may also help to prevent declining response rates by making it easier, more intuitive and more enjoyable for the respondent to participate in online surveys.

In our opinion, future marketing research will be more integrated with other information tools. For example, customer relationship management software could include online surveys and research reporting applications. Also decision support tools such as dashboards can integrate online surveys to receive continuous input from customers to predict future behavior as well as revenues.

APPENDIX

The Survey Participation Inventory

I participated in this online survey

Interest

- (1) ...because I find the subject of this online survey interesting
 - (2) ...because I find the topic of this survey exciting
 - (3) ...because I believe the topic is of significant importance
 - (4) ...because I am interested in the latest developments in marketing
-

Give Opinion

- (1) ...because I find it important to voice my opinion
 - (2) ...because I want to share my point of view in this online survey
 - (3) ...because I appreciate that I am asked to express my opinion
 - (4) ...because it matters to me that I can communicate my point of view
-

Curiosity

- (1) ...because I want to learn something new
 - (2) ...because I like the idea of learning about the topic at hand
 - (3) ...because I am fond of exploring new ideas
 - (4) ...because I was curious to find out more about this topic
-

Enjoyment

- (1) ...because it is fun to do
 - (2) ...because I find participating in this online survey appealing
 - (3) ...because I think that participating is quite enjoyable
 - (4) ...because I think it is pleasurable
-

Helping

- (1) ...because I want to help companies
 - (2) ...because I believe it is important that I assist researchers and companies in collecting data
 - (3) ...because the thought of actively helping out companies on how to improve their products and services stimulates me to participate
 - (4) ...because I find it is my responsibility to support research in general
-

Recognition

- (1) ...because I have the feeling that I receive acknowledgement for participating
 - (2) ...because I am concerned about what the researcher would think about me if I do not participate
 - (3) ...because I like the thought that others recognize what a dedicated panel member I am
 - (4) ...because I like the notion that others acknowledge my efforts in filling out surveys
-

Incentives

- (1) ...because of the incentive or prize I can win in return for my participation
 - (2) ...because I calculated that the reward would compensate my effort
 - (3) ...because I support a charity goal in return for my participation
 - (4) ...because the reward is important to me
-

Obligation

- (1) ...because I feel that it is my duty to participate
 - (2) ...because I feel obliged to participate since I am a registered panel member
 - (3) ...because researchers and companies should be able to count on my support
 - (4) ...because I feel compelled to participate
-

Scale items for Effort, Response Quality and Topic Salience

Effort

- (1) I put a lot of effort into filling out this online questionnaire
 - (2) I tried very hard to fill out this survey
 - (3) I concentrated a lot while participating in this online research
 - (4) I was careful about which answer I have chosen for every question
-

Response quality

- (1) I would judge my performance in filling out this questionnaire as high
 - (2) I have the feeling I could answer all questions well
 - (3) The quality of my answers is very high
 - (4) Researchers can use my answers to draw important conclusions
-

Topic Salience (Control variable)

- (1) I have great interest in this topic
 - (2) This topic is fascinating
 - (3) I like this topic
 - (4) The topic interests me
-

Below are eight reasons for survey participation. Please allocate 100 points among the reasons so that your allocation reflects the relative importance you attach to each reason. The more points a reason receives, the more important the reason is. If a reason is not at all important, assign 0 points.

- (1) Give opinion
 - (2) Curiosity
 - (3) Incentives
 - (4) Interest
 - (5) Helping
 - (6) Recognition
 - (7) Enjoyment
 - (8) Obligation
-

REFERENCES

A

Agresti, Allan (1990), *Categorical Data Analysis*, New York: Wiley.

Albaum, Gerald S., Felicitas Evangelista, and Nila Medina (1998), "Role of Response Behavior Theory in Survey Research: A Cross-National Study," *Journal of Business Research*, 42 (2), 115-25.

Alexandris, Konstantinos, Charalambos Tsorbatzoudis, and George Grouios (2002), "Perceived Constraints on Recreational Sport Participation: Investigating their Relationship with Intrinsic Motivation, Extrinsic Motivation and Amotivation," *Journal of Leisure Research*, 34 (3), 233-52.

Amabile, Teresa M., Karl G. Hill, Beth A. Hennessey, and Elizabeth M. Tighe (1994), "The Work Preference Inventory: Assessing Intrinsic and Extrinsic Motivational Orientations," *Journal of Personality and Social Psychology*, 66 (5), 950-67.

American Marketing Association retrieved May 15, 2005, from: <http://www.marketingpower.com/mg-dictionary-view1884.php>

Anderson, James C. and David W. Gerbing (1988), "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin*, 103 (3), 411-23.

Areni, C., M. Ferrell, and J. Wilcox (1999), "The Effects of Need for Cognition and Topic Importance on the Latency and Extremity of Responses to Attitudinal Inquiries," in *Advances in Consumer Research*, Provo, Utah,

Aspirations Index retrieved February 5, 2004, from: <http://www.psych.rochester.edu/SDT/measures/aspir.html>.

B

Bagozzi, Richard P., Claes Fornell, and David F. Larcker (1981), "Canonical Correlation Analysis as a Special Case of a Structural Relations Model," *Multivariate Behavioral Research*, 16 (4), 437-54.

Bagozzi, Richard P. and Paul R. Warshaw (1990), "Trying to Consume," *Journal of Consumer Research*, 17 (2), 127-40.

Baumgartner, Hans and Jan-Benedict E.M. Steenkamp (2001), "Response Styles in Marketing Research: A Cross-National Investigation," *Journal of Marketing Research*, 38 (2), 143-56.

Bettencourt, Lance A. and Stephen W. Brown (1997), "Contact Employees: Relationship among Work Place Fairness, Job Satisfaction, and Prosocial Service Behaviors," *Journal of Retailing*, 73 (1), 39-61.

- Bickart, Barbara and David Schmittlein (1999), "The Distribution of Survey Contact and Participation in the United States: Constructing a Survey-Based Estimate," *Journal of Marketing Research*, 36 (2), 286-94.
- Birgelen, Marcel van, Ko de Ruyter, Ad de Jong, and Martin G.M. Wetzels (2002), "Customer Evaluations of After-Sales Service Contact Models: An Empirical Analysis of National Culture's Consequences," *International Journal of Research in Marketing*, 19 (1), 43-64.
- Bollen, Kenneth A. (1989), *Structural Equations with Latent Variables*, New York: Wiley.
- Bosnjak, Michael, Tracy L. Tuten, and Werner W. Wittmann (2005), "Unit (Non)Response in Web-Based Access Panel Surveys: An Extended Planned Behavior Approach," *Psychology & Marketing*, 22 (6), 489-505.
- Brennan, Mike and Janet Hoek (1992), "The Behavior of Respondents, Nonrespondents, and Refusers Across Mail Surveys," *Public Opinion Quarterly*, 56, 530-35.
- Brennan, Robert L. (2001), *Generalizability Theory*, New York: Springer-Verlag.
- Brown, Tom J., Gilbert A. Churchill Jr., and J. Paul Peter (1993), "Improving the Measurement of Service Quality," *Journal of Retailing*, 69 (1), 127-39.
- Browne, Michael W. (1984), "Asymptotically Distribution-Free Methods for the Analysis of Covariance Structures," *British Journal of Mathematical and Statistical Psychology*, 37, 62-83.
- Buchanan, Tom and John L. Smith (1999), "Using the Internet for Psychological Research: Personality Testing on the World Wide Web," *British Journal of Psychology*, 90 (1), 125-44.
- Burke, Raymond R., Arvind Rangaswamy, and Sunil Gupta (1999), "Rethinking Marketing Research in the Digital World," retrieved March 25, 2002, from: http://commerce.concordia.ca/gkersten/e-business_new/emarketing_res.pdf.

C

- Cacioppo, John T., Richard E. Petty, and Chuan Feng Kao (1984), "The Efficient Assessment of Need for Cognition," *Journal of Personality Assessment*, 48 (3), 306-07.
- Calinski, T. and J. Harabasz (1974), "A Dendrite Method for Cluster Analysis," *Communications in Statistics*, 3 (1), 1-27.
- Chou, Chih-Ping, Peter M. Bentler, and Albert Satorra (1991), "Scaled Test Statistics and Robust Standard Errors for Nonnormal Data in Covariance Structure Analysis: A Monte Carlo Study," *British Journal of Mathematical and Statistical Psychology*, 44, 347-57.

Church, Allan H. (1993), "Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis," *Public Opinion Quarterly*, 57 (1), 62-79.

Churchill, Gilbert A. Jr. (1979), "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research*, 16 (1), 64-73.

Cobanoglu, Cihan, Bill Warde, and Patrick J. Moreo (2001), "A Comparison of Mail, Fax and Web-based Survey Methods," *International Journal of Market Research*, 43 (4), 441-52.

comScore networks (2005), "The Professional Respondent Problem In Online Survey Panels Today," in *MRA Annual Conference*,

Cook, Colleen, Fred Heath, and Russell Thompson (2000), "A Meta-Analysis of Response Rates in Web- or Internet-based Surveys," *Educational & Psychological Measurement*, 60 (6), 821-36.

Couper, Mick P. (2000), "Web Surveys: A Review of Issues and Approaches," *Public Opinion Quarterly*, 64 (4), 464-94.

Couper, Mick P., Michael W. Traugott, and Mark J. Lamias (2001), "Web Survey Design and Administration," *Public Opinion Quarterly*, 65 (2), 230-53.

Cronbach, Lee J., Goldine C. Gleser, Harinder Nanda, and Nageswari Rajaratnam (1972), *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*, New York: Wiley.

Cronbach, Lee J., Nageswari Rajaratnam, and Goldine C. Gleser (1963), "Theory of Generalizability: A Liberalization of Reliability Theory," *British Journal of Statistical Psychology*, 16 (2), 137-63.

Crowne, Douglas P. and David Marlowe (1960), "A New Scale of Social Desirability Independent of Psychopathology," *Journal of Consulting Psychology*, 24 (4), 349-54.

Csikszentmihalyi, Mihaly (1975), *Beyond Boredom and Anxiety*, New York: Jossey-Bass.

D

Daugherty, Terry, Wei-Na Lee, Harsha Gangadharbatla, Kihan Kim, and Sounthaly Outhavong (2005), "Organizational Virtual Communities: Exploring Motivations Behind Online Panel Participation," *Journal of Computer-Mediated Communication*, 10 (4).

Day, George S. (1994), "the Capabilities of Market-Driven Organizations," *Journal of Marketing*, 58 (October), 37-52.

Deci, Edward L. and Richard M. Ryan (1985), *Intrinsic Motivation and Self-Determination in Human Behavior*, New York: Plenum Press.

- Deutsdens, Elisabeth C., Ko de Ruyter, Martin G.M. Wetzels, and Paul Oosterveld (2004), "Response Rate and Response Quality of Internet-based Surveys: An Experimental Study," *Marketing Letters*, 15 (1), 21-36.
- Diamond, William D. and Rajiv K. Kashyap (1997), "Extending Models of Prosocial Behavior to Explain University Alumni Contributions," *Journal of Applied Social Psychology*, 27 (10), 915-28.
- Dillman, Don A. (2000), *Mail and Internet Surveys, The Tailored Design Method*, 2nd ed. New York: Wiley.
- Dillman, Don A. and Dennis K. Bowker (2001), "The Web Questionnaire Challenge to Survey Methodologists," in *Dimensions of Internet Science*, Ulf-Dietrich Reips & Michael Bosnjak, ed. Lengerich: Pabst Science Publishers.
- Dillman, Don A., Glenn Phelps, Robert D. Tortora, Karen Swift, Karen Kohrell, and Jodi Berck (2001), "Response Rate and Measurement Differences in Mixed Mode Surveys Using Mail, Telephone, Interactive Voice Response and the Internet," retrieved May 14, 2003, from: http://survey.sesrc.wsu.edu/dillman/papers/Mixed%20Mode%20ppr%20_with%20Gallup_%20POQ.pdf.
- Dillman, Don A., Robert D. Tortora, John Conradt, and Dennis Bowker (1998), "Influence of Plain vs. Fancy Design on Response Rates for Web Surveys," in *Joint Statistical Meetings*, Dallas, Texas,
- Duda, Richard O. and Peter E. Hart (1973), *Pattern Classification and Scene Analysis*, New York: Wiley.

E

- Efron, Bradley and Robert Tibshirani (1986), "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy," *Statistical Science*, 1 (1), 54-77.
- Epstein, Joel, W. Dean Klinkenberg, D. Wiley, and L. McKinley (2001), "Insuring Sample Equivalence across Internet and Paper-and-Pencil Assessments," *Computers in Human Behavior*, 17 (3), 339-46.
- ESOMAR (2004a), "Marketing and Opinion Research Industry in the USA is the Largest Market Worldwide," retrieved October 14, 2004, from: <http://www.esomar.org/esomar/show/id=107403>.
- ESOMAR (2004b), "MR Expenditure Worldwide Up By 5.1% in 2003," retrieved April 1, 2005, from: <http://www.esomar.org/web/show/id=148285>.
- Evangelista, Felicitas, Gerald S. Albaum, and Patrick Poon (1999), "An Empirical Test of Alternative Theories of Survey Response Behavior," *Journal of the Market Research Society*, 41 (2), 227-44.
- Everitt, Brian S. (1992), *The Analysis of Contingency Tables*, London: Chapman and Hall/CRC.

F

Feldman, Jack M. and John G. Lynch Jr. (1988), "Self-Generated Validity and Other Effects of Measurement of Belief, Attitude, Intention, and Behavior," *Journal of Applied Psychology*, 73 (3), 421-35.

Finn, Adam and Ujwal Kayandé (1997), "Reliability Assessment and Optimization of Marketing Measurement.," *Journal of Marketing Research*, 34 (2), 262-75.

Fornell, Claes and David F. Larcker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18 (3), 39-50.

Forrest, Ed (1999), *Internet Marketing Research, Resources and Techniques*, Sydney: Irwin/McGraw-Hill.

Fox, Richard J., Melvin R. Crask, and Jonghoon Kim (1988), "Mail Survey Response Rates," *Public Opinion Quarterly*, 52 (4), 467-91.

Furse, David H. and David W. Stewart (1982), "Monetary Incentives versus Promised Contribution to Charity: New Evidence on Mail Survey Response," *Journal of Marketing Research*, 19 (3), 375-80.

G

Games, Paul A. and James F. Howell (1976), "Pairwise Multiple Comparison Procedures with Unequal N's and/or Variances: a Monte Carlo Study," *Journal of Educational Statistics*, 1 (2), 113-25.

Goetz, Edward G., Tom R. Tyler, and Fay Lomax Cook (1984), "Promised Incentives in Media Research: A Look at Data Quality, Sample Representativeness, and Response Rate," *Journal of Marketing Research*, 21 (2), 148-54.

Göritz, Anja S. (2004), "The Impact of Material Incentives on Response Quantity, Response Quality, Sample Composition, Survey Outcome, and Cost in Online Access Panels," *International Journal of Market Research*, 46 (3), 327-45.

Göritz, Anja S., Nicole Reinhold, and Bernad Batinic (2002), "Online Panels," in *Online Social Sciences*, Bernad Batinic and Ulf-Dietrich Reips, eds. Seattle: WA Hogrefe & Huber, 27-47.

Gouldner, Alvin W. (1960), "The Norm of Reciprocity: A Preliminary Statement," *American Sociological Review*, 25 (2), 161-78.

Goyder, John (1987), *The Silent Minority: Nonrespondents on Sample Surveys*, Cambridge, UK: Polity Press.

Grandcolas, Ursula, Ruth Rettie, and Kira Marusenko (2003), "Web Survey Bias: Sample or Mode Effect?," *Journal of Marketing Management*, 19 (5/6), 541-61.

- Green, Paul E., Richard M. Johnson, and William D. Neal (2003), "The *Journal of Marketing Research*: Its Initiation, Growth, and Knowledge Dissemination," *Journal of Marketing Research*, 40 (1), 1-9.
- Greenbaum, Thomas (1998), "Internet Focus Groups are not Focus Groups - So Don't Call Them That," retrieved February 26, 2002, from: http://www.quirks.com/articles/article.asp?arg_ArticleId=355.
- Greenleaf, Eric A. (1992a), "Improving rating Scale Measures by Detecting and Correcting Bias Components in Some Response Styles," *Journal of Marketing Research*, 29 (2), 176-88.
- Greenleaf, Eric A. (1992b), "Measuring Extreme Response Style," *Public Opinion Quarterly*, 56 (3), 328-51.
- Greenspan, Robyn (2002), "E-Shopping Around the World," retrieved August 20th, 2002, from: http://cyberatlas.internet.com/markets/retailing/article/0,,6061_1431461,00.html.
- Griffis, Stanley E., Thomas J. Goldsby, and Martha Cooper (2003), "Web-Based and Mail Surveys: A Comparison of Response, Data, and Cost," *Journal of Business Logistics*, 24 (2), 237-58.
- Groves, Robert M., Robert B. Cialdini, and Mick P. Couper (1992), "Understanding the Decision to Participate in a Survey," *Public Opinion Quarterly*, 56 (4), 475-795.
- Groves, Robert M., Stanley Presser, and Sarah Dipko (2004), "The Role of Topic Interest in Survey Participation Decisions," *Public Opinion Quarterly*, 68 (1), 2-31.
- Groves, Robert M., Eleanor Singer, and Amy Corning (2000), "Leverage-Saliency Theory of Survey Participation," *Public Opinion Quarterly*, 64 (3), 299-308.

H

- Hansen, Robert A. (1980), "A Self-Perception Interpretation of the Effect of Monetary and Nonmonetary Incentives on Mail Survey Respondent Behavior," *Journal of Marketing Research*, 17 (1), 77-83.
- Hartley, H. O., J. N. K. Rao, and L. R. LaMotte (1978), "A Simple Synthesis-Based Method of Variance Component Estimation," *Biometrics*, 34, 233-42.
- Heberlein, Thomas A. and Robert Baumgartner (1978), "Factors Affecting Response Rates to Mailed Questionnaires: a Quantitative Analysis of the Published Literature," *American Sociology Review*, 43 (4), 447-62.
- Helgeson, James G., Kevin E. Voss, and Willbann D. Terpening (2002), "Determinants of Mail Survey Response: Survey Design Factors and Respondent Factors," *Psychology and Marketing*, 19 (3), 303-28.

Hogg, Allen (2002), "Conducting Online Research," retrieved June 18, 2003, from: <http://www.burke.com/whitepapers/PDF/B.WhitePaperVol3-2002-Iss2.pdf>.

Holbrook, Morris B. (1986), "Aims, Concepts, and Methods for the Representation of Individual Differences in Esthetic Responses to Design Features," *Journal of Consumer Research*, 13 (3), 337-47.

<http://www.bloomerce.com/>.

<http://www.misterpoll.com/>.

"<http://www.opinionworld.co.uk/>."

Hunt, J. McV. (1965), "Intrinsic Motivation and Its Role in Psychological Development," in *Nebraska Symposium on Motivation*, D. Levine, ed. Vol. 13. Lincoln: University of Nebraska Press, 189-282.

Hupprich, Laura and Maria Bumatay (2002), "Hong Kong Leads the World in High-Speed Internet Connections, According to Nielsen/Netratings' Global Internet Trend Survey," retrieved May 27, 2003, from: http://www.nielsen-netratings.com/pr/pr_020815.pdf.

I

Illieva, Janet, Steve Baron, and Nigel M. Healey (2002), "Online Surveys in Marketing Research: Pros and Cons," *International Journal of Market Research*, 44 (3), 361-82.

Intrinsic Motivation Inventory (IMI) retrieved February 5, 2004, from: <http://www.psych.rochester.edu/SDT/measures/intrins.html>.

Iqbal, Zafar, Rohit Verma, and Roger Baran (2003), "Understanding Consumer Choices and Preferences in Transaction-Based e-Services," *Journal of Service Research*, 6 (1), 51-65.

Irons, Karl (2000), "Moving Marketing Research Activities to the Internet," in *Marketing Research in a .com Environment*, Richard Brookers, ed. Amsterdam: ESOMAR.

Izard, Carroll E. (1977), *Human Emotions*, New York: Plenum Press.

J

Johnson, Eric J. (2001), "Digitizing Consumer Research," *Journal of Consumer Research*, 28 (2), 331-36.

Jones, Wesley H. and Gerald Linda (1978), "Multiple Criteria Effects in a Mail Survey Experiment," *Journal of Marketing Research*, 15 (May), 280-84.

Jöreskog, Karl G. (1971), "Simultaneous Factor Analysis in Several Populations," *Psychometrika*, 36 (4), 409-26.

K

Kane, Michael T., Gerald M. Gillmore, and Terrence J. Crooks (1976), "Student Evaluations of Teaching: The Generalizability of Class Means," *Journal of Educational Measurement*, 13 (3), 171-84.

Kanuk, Leslie and Conrad Berenson (1975), "Mail Survey and Response Rates: A Literature Review," *Journal of Marketing Research*, 12 (4), 440-53.

Kenhove, Patrick van, Katrien Wijnen, and Kristof De Wulf (2002), "The Influence of Topic Involvement on Mail-Survey Response Behavior," *Psychology & Marketing*, 19 (3), 293-301.

Kiesler, Sara and Lee S. Sproull (1986), "Response Effects in the Electronic Survey," *Public Opinion Quarterly*, 50 (3), 402-13.

Kirlin, Mary K. (2002), "Understanding How Organizations Affect the Civic Engagement of Adolescent Participants," *New Directions for Philanthropic Fundraising*, 2002 (38), 19-35.

Klassen, Robert D. and Jennifer Jacobs (2001), "Experimental Comparison of Web, Electronic and Mail Survey Technologies in Operations Management," *Journal of Operations Management*, 19 (6), 713-28.

Knapp, Herschel and Stuart A. Kirk (2003), "Using Pencil and Paper, Internet and Touch-Tone Phones for Self-Administered Surveys: Does Methodology Matter?," *Computers in Human Behavior*, 19 (1), 117-34.

Kohli, Ajay K. and Bernard J. Jaworski (1990), "Market Orientation: The Construct, Research Propositions, and Managerial Implications," *Journal of Marketing*, 54 (April), 1-18.

Krosnick, Jon A. (1991), "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys," *Applied Cognitive Psychology*, 5, 213-36.

Krosnick, Jon A. (1999), "Survey Research," *Annual Review of Psychology*, 50 (1), 537-67.

L

Lambert, Zarrel V. and Richard M. Durand (1975), "Some Precautions in Using Canonical Analysis," *Journal of Marketing Research*, 12 (4), 468-75.

Linsky, Arnold S. (1975), "Stimulating Responses to Mailed Questionnaires," *Public Opinion Quarterly*, 39 (1), 82-101.

Lozar Manfreda, Katja (2001), "Web Survey Errors," Dissertation, University of Ljubljana.

Lozar Manfreda, Katja, Zenel Batagelj, and Vasja Vehovar (2002), "Design of Web Survey Questionnaires: Three Basic Experiments," retrieved January 20, 2003, from: <http://www.ascusc.org/jcmc/vol7/issue3/vehovar.html>.

Lynch, Jr., John G. and Dan. Ariely (2000), "Wine Online: Search Costs Affect Competition on Price, Quality, and Distribution," *Marketing Science*, 19 (1), 83-103.

M

Malhotra, Naresh K. (1999), *Marketing Research: An Applied Orientation*, 3rd ed. New Jersey: Prentice Hall.

Mandel, Naomi and Eric J. Johnson (2002), "When Web Pages Influence Choice: Effects of Visual Primes on Experts and Novices," *Journal of Consumer Research*, 29 (2), 235-45.

Mardia, Kanti V. (1970), "Measures of Multivariate Skewness and Kurtosis with Applications," *Biometrika*, 57, 519-30.

McDaniel, Stephen W. and C.P. Rao (1980), "The Effect of Monetary Inducement on Mailed Questionnaire Response Quality," *Journal of Marketing Research*, 17 (2), 265-68.

McDonald, Heath and Stewart Adam (2003), "A Comparison of Online and Postal Data Collection Methods in Marketing Research," *Marketing Intelligence & Planning*, 21 (2), 85-95.

Metzke, Nicolas and Katherine Allan (2005), "Researching the Researchers: The Online Research Barometer," in *ESOMAR Panel Research Conference*, Budapest, April 2005.

Meuter, Matthew L., Amy L. Ostrom, Robert I. Roundtree, and Mary Jo Bitner (2000), "Self-Service Technologies: Understanding Customer Satisfaction with Technology-Based Service Encounters," *Journal of Marketing*, 64 (3), 50-64.

Miller, Thomas W. and Peter R. Dickson (2001), "Online Market Research," *International Journal of Electronic Commerce*, 5 (3), 139-67.

Milligan, Glenn W. and Martha C. Cooper (1985), "An Examination of Procedures for Determining the Number of Clusters in a Data Set," *Psychometrika*, 50 (1), 159-79.

Milligan, Glenn W. and Martha C. Cooper (1986), "A Study of the Comparability of External Criteria for Hierarchical Cluster Analysis," *Multivariate Behavioral Research*, 21 (4), 441-58.

Montoya-Weiss, Mitzi M., A.P. Massey, and D.L. Clapper (1998), "On-line Focus Groups: Conceptual Issues and a Research Tool," *European Journal of Marketing*, 32 (7/8), 713-23.

Moorman, Christine and Roland T. Rust (1999), "The Role of Marketing," *Journal of Marketing*, 63 (Special Issue 1999), 180-97.

Morin, Richard (October 28, 2004), "Don't Ask Me: As Fewer Cooperate on Polls, Criticism and Questions Mount", in *Washingtonpost.com*

N

Narver, John C. and Stanley F. Slater (1990), "The Effect of Marketing Orientation on Business Profitability," *Journal of Marketing*, 54 (October), 20-35.

Nederhof, Anton J. (1983), "The Effects of Material Incentives in Mail Surveys: Two Studies," *Public Opinion Quarterly*, 47 (1), 107-11.

Nielsen, Jakob (2000), *Designing Web Usability*, Indianapolis: New Riders Publishing.

Novak, Thomas P., Donna L. Hoffman, and Yiu-Fai Yung (2000), "Measuring the Customer Experience in Online Environments: A Structural Modeling Approach," *Marketing Science*, 19 (1), 22-42.

NUA Internet Surveys (2002), "Ipsos-Reid: Internet Use Climbing in Most Markets," retrieved January 29, 2004, from: http://www.nua.com/surveys/index.cgi?f=VS&art_id=905358657&rel=true.

NUA Internet Surveys (2003), "Nielsen Netratings: Global Net Population Increases," retrieved May 27, 2003, from: http://www.nua.ie/surveys/index.cgi?f=VS&art_id=905358729&rel=true.

Nunnally, J.C. and I.H. Bernstein (1994), *Psychometric Theory*, New York: McGraw-Hill.

O

OnlineSurveyServices "Cost Comparison of Paper Survey vs. Online Survey," retrieved December 8, 2003, from: <http://www.onlinesurveyservices.com/olss/cost.htm>.

P

Parasuraman, A., Valarie A. Zeithaml, and Leonard L. Berry (1988), "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality," *Journal of Retailing*, 64 (1), 12-40.

Paulhus, Delroy L. (1991), "Measurement and Control of Response Bias," in *Measures of Personality and Social Psychological Attitudes*, John P. Robinson and Phillip R. Shaver and Lawrence S. Wright, eds. San Diego, CA: Academic Press, 17-59.

Peter, J.Paul (1979), "Reliability: A Review of Psychometric basic and Recent Marketing Practices," *Journal of Marketing Research*, 16 (1), 6-17.

- Petty, Richard E. and John T. Cacioppo (1986), *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*, New York: Springer-Verlag.
- Pierce, W. David, Judy Cameron, Katherine M. Banko, and Sylvia So (2003), "Positive Effects of Rewards and Performance Standards on Intrinsic Motivation," *The Psychological Record*, 53 (4), 561-79.
- "Population Explosion," retrieved May 16th, 2005, from: <http://www.clickz.com/stats/sectors/geographics/article.php/151151>.
- Porst, Rolf and Christa Von Briel (1995), "Wären Sie Vielleicht Bereit, Sich Gegebenenfalls Noch Einmal Befragen zu Lassen? Oder: Gründe für die Teilnahme an Panelbefragungen," *ZUMA-Arbeitsbericht Nr. 95/04*.
- Posdakoff, Philip M. and Scott B. Mackenzie (1994), "Organizational Citizenship Behaviors and Sales Unit Effectiveness," *Journal of Marketing Research*, 31 (3), 351-63.
- Pretty, Grace H. and C. Seligman (1983), "Affect and the Overjustification Effect," *Journal of Personality and Social Psychology*, 46, 1241-53.
- "Privacy and Electronic Communications (EC Directive)," retrieved September 23, 2005, from: <http://www.informationcommissioner.gov.uk/eventual.aspx?id=35>.
- Punj, Girish and David W. Stewart (1983), "Cluster Analysis in Marketing Research: Review and Suggestions for Application," *Journal of Marketing Research*, 20 (2), 134-48.

R

- Rentz, Joseph O. (1987), "Generalizability Theory: A Comprehensive Method for Assessing and Improving the Dependability of Marketing Measures," *Journal of Marketing Research*, 24 (1), 19-28.
- Rentz, Joseph O. (1988), "An Exploratory Study of the Generalizability of Selected Marketing Measures," *Academy of Marketing Science*, 16 (1), 141-50.
- Roberson, M.T. and E. Sundstrom (1990), "Survey Design, Return Rates, and Response Favorableness in an Employee Attitude Survey," *Journal of Applied Psychology*, 75, 354-57.
- Robinson, John P., Alan. Neustadt, and Meyer Kestnbaum (2002), "Why Internet Samples are Inherently Biased," in *The American Association for Public Opinion Research (AAPOR) 57th Annual Conference*, St. Pete Beach, Florida,
- Rosenblum, Jeff (2001), "Give and Take," retrieved June 14, 2002, from: http://www.quirks.com/articles/article_print.asp?arg_articleid=705.
- Roster, Catherine A., Robert D. Rogers, Gerald Alba, and Darin Klein (2004), "A Comparison of Response Characteristic from Web and Telephone Surveys," *International Journal of Market Research*, 46 (3), 359-73.

Ruyter, Ko de and Martin G.M. Wetzels (2000), "With a Little Help from my Fans - Extending Models of Pro-Social Behavior to Explain Supporters' Intentions to Buy Soccer Club Shares," *Journal of Economic Psychology*, 21 (4), 387-409.

Rust, Roland T. and Bruce Cooil (1994), "Reliability Measures for Qualitative Data: Theory and Implications," *Journal of Marketing Research*, 31 (1), 1-14.

Ryan, Richard M. and Edward L. Deci (2000), "Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions," *Contemporary Educational Psychology*, 25 (1), 54-67.

S

Salkin, Harvey M. (1975), *Integer programming*, Massachusetts: Addison-Wesley.

Sanders, P. F., T. J. J. M. Theunissen, and S. M. Baas (1989), "Minimizing the Number of Observations: A Generalization of the Spearman-Brown Formula," *Psychometrika*, 54, 587-98.

Satorra, Albert and Peter M. Bentler (1994), "Corrections to Test Statistic and Standard Errors in Covariance Structure Analysis," in *Analysis of Latent Variables in Developmental Research*, A. von Eye and C. C. Clogg, eds. Newbury Park, CA: Sage Publications, 399-419.

Satorra, Albert and Peter M. Bentler (2001), "A Scaled Difference Chi-Square Test Statistic for Moment Structure Analysis," *Psychometrika*, 66 (4), 507-14.

Schaefer, David R. and Don A. Dillman (1998), "Development of a Standard E-Mail Methodology," *Public Opinion Quarterly*, 62 (3), 378-97.

Schaninger, Charles M. and W. Christian Buss (1986), "Removing Response-Style Effects in Attribute-Determinance Ratings to Identify Market Segments," *Journal of Business Research*, 14 (3), 237-52.

Schillewaert, Niels and Pascale Meulemeester (2005), "Comparing Response Distributions of offline and Online Data Collection Methods," *International Journal of Market Research*, 47 (2), 163-78.

Scholl, N., S. Mulders, and R. Drent (2002), "On-line Qualitative Market Research: Interviewing the World at a Finger Tip," *Qualitative Market Research: An International Journal*, 5 (3), 210-23.

Schuldt, Barbara A. and Jeff W. Totten (1994), "Electronic Mail vs. Mail Survey Response Rates," *Marketing Research*, 6 (1), 36-39.

Searle, Shayle R., George Casella, and Charles E. McCulloch (1992), *Variance Components*, New York: Wiley.

Sharma, Subhash and Danny Weathers (2003), "Assessing Generalizability of Scales Used in Cross-National Research," *International Journal of Research in Marketing*, 20 (3), 287-95.

- Sheehan, Kim B. (2001), "E-mail Survey Response Rates: A Review," retrieved January 15, 2003, from: <http://www.ascusc.org/jcmc/vol6/issue2/sheehan.html>.
- Sheehan, Kim B. and S.J. McMillan (1999), "Response Variation in E-Mail Surveys: An Exploration," *Journal of Advertising Research*, 39 (4), 45.
- Shermis, Mark D. and Danielle Lombard (1999), "A Comparison of Survey Data Collected by Regular Mail and Electronic Mail Questionnaires," *Journal of Business & Psychology*, 14 (2), 341-54.
- Shulman, Art (1973), "A Comparison of Two Scales on Extremity Response Bias," *Public Opinion Quarterly*, 37 (3), 407-12.
- Simsek, Zeki and John F. Veiga (2001), "A Primer on Internet Organizational Surveys," *Organizational Research Methods*, 4 (3), 218-35.
- Singer, Eleanor and Richard A. Kulka (2000), "Paying Respondents for Survey Participation," in Survey Methodology Program, Survey Research Center (SRC) at the Institute for Social Research, University of Michigan.
- Smither, James W., Alan G. Walker, and Michael K.T. Yap (2004), "An Examination of the Equivalence of Web-Based Versus Paper-and-Pencil Upward Feedback Ratings: Rater- and Ratee-Level Analyses," *Education and Psychological Measurement*, 64 (1), 40-61.
- Solomon, David J. (2001), "Conducting Web-Based Surveys," retrieved January, 9th, 2002, from: <http://ericae.net/pare/getvn.asp?v=7&n=19>.
- Sparrow, Nick and John Curtice (2004), "Measuring the Attitudes of the General Public via Internet Polls: An Evaluation," *International Journal of Market Research*, 46 (1), 23-44.
- Sproull, Lee and Sara Kiesler (1986), "Reducing Social Context Cues: Electronic Mail in Organizational Communication," *Management Science*, 32 (11), 1492.
- Srinivasan, Narasimhan and Brian T. Ratchford (1991), "An Empirical Test of a Model of External Search for Automobiles," *Journal of Consumer Research*, 18 (2), 233-42.
- Srivastava, Muni S. (1984), "Measure of Skewness and Kurtosis and a Graphical Method for Assessing Multivariate Normality," *Statistical Probability Letters*, 2 (5), 263-367.
- Stanton, Jeffrey M. (1998), "An Empirical Assessment of Data Collection Using the Internet," *Personnel Psychology*, 51 (3), 709-25.
- Steenkamp, Jan-Benedict E.M. and Hans Baumgartner (1998), "Assessing Measurement Invariance in Cross-National Consumer Research," *Journal of Consumer Research*, 25 (1), 78-90.

T

Taylor, Humphrey (2000), "Does Internet Research Work? Comparing Online Survey Results with Telephone Survey," *International Journal of Market Research*, 42 (1), 51-63.

Thompson, Bruce (1984), *Canonical Correlation Analysis: Uses and Interpretation*, Thousand Oaks, CA.

Timmerman, A. (2002), "Introduction to the Application of Web-Based Surveys," retrieved August 5, 2005, from: http://tps.dpi.state.nc.us/ncd/surveyweb/introduction_to_the_application.html.

Toubia, Oliver, John R. Hauser, and Duncan I. Simester (2004), "Polyhedral Methods for Adaptive Choice-Based Conjoint Analysis," *Journal of Marketing Research*, 41 (1), 116-31.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski (2000), *The Psychology of Survey Response*: Cambridge University Press.

Truell, Allen D., James E. Bartlett II, and Melody W. Alexander (2002), "Response Rate, Speed, and Completeness: A Comparison of Internet-Based and Mail Surveys," *Behavior Research Methods, Instruments, & Computers*, 34 (1), 46-49.

V

Vandenberg, Robert J. (2002), "Toward a Further Understanding of and Improvement in Measurement Invariance Methods and Procedures," *Organizational Research Methods*, 5 (2), 139-58.

Vandenberg, Robert J. and Charles E. Lance (2000), "A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research," *Organizational Research Methods*, 3 (1), 4-70.

Vehovar, Vasja, Katja Lozar Manfreda, and Zenel Batagelj (1999), "Web Surveys: Can the Weighting Solve The Problem?," in *Proceedings of the Survey Research Methods Section*

W

Warriner, Keith, John Goyder, Heidi Gjertsen, Paula Hohner, and Kathleen McSpurren (1996), "Charities, No; Lotteries, No; Cash, Yes," *Public Opinion Quarterly*, 60 (4), 542-62.

Warwick, Donald P. and Charles Andrew Lininger (1975), *The Sample Survey: Theory and Practice*, New York: McGraw-Hill.

Weijters, Bert, Niels Schillewaert, and Maggie Geuens (2004), "Measurement Bias due to Response Styles: A Structural Equation Model Assessing the Effect of Modes of Data-Collection," in Vlerick Leuven Gent Working Paper Series 2004/20.

West, Stephan G., John F. Finch, and Patrick J. Curran (1995), "Structural Equation Models with Nonnormal Variables," in *Structural Equation Modeling: Concepts, Issues and Applications*, R.H. Hoyle, ed. Thousand Oaks, CA: Sage Publications, 56-75.

Wind, Jerry (1997), "Start Your Engines: Gear up for Challenges Ahead with Innovative Marketing Research Products and Services," *Marketing Research Forum*, Winter, 4-11.

Winer, Russell S. (1983), "Attrition Bias in Econometric Models Estimated With Panel Data," *Journal of Marketing Research*, 20 (2), 177-86.

Wiseman, Frederick. (1972), "Methodological Bias in Public Opinion Surveys," *Public Opinion Quarterly*, 36 (1), 105-08.

Y

Yammarino, Francis J., Steven J. Skinner, and Terry L. Childers (1991), "Understanding Mail Survey Response Behavior," *Public Opinion Quarterly*, 55 (4), 613-39.

Yoon, Mahn Hee and Jaebeom Suh (2003), "Organizational Citizenship Behaviors and Service Quality as External Effectiveness of Contact Employees," *Journal of Business Research*, 56 (8), 597-611.

Yu, Julie and Harris Cooper (1983), "A Quantitative Review of Research Design Effects on Response Rates to Questionnaires," *Journal of Marketing Research*, 20 (1), 36-44.

Z

Zeithaml, Valarie A. and Mary Jo Bitner (2003), *Services Marketing: Integrating Customer Focus Across the Firm*, 3rd ed. New York: McGraw-Hill.

Curriculum Vitae

Elisabeth C. Deutskens was born on the 7th of August 1977 in Marl, Germany. She attended the Lise-Meitner Gymnasium in Geldern and graduated in 1997. After working for McDonald's for several years, she was promoted to an assistant of the store manager and received training in the area of employee leadership and communication. Her duties included first-line management as well as hiring, training and education of employees. In September 1997, she started studying International Management at the Faculty of Economics and Business Administration of Maastricht University. She received her Master of Economics in September 2001. During her studies, she spent one semester at the University of California – Santa Barbara. In addition, she gathered her first teaching experiences as a student assistant teacher for several organization and marketing courses. She joined the Department of Marketing as a Ph.D. Student in September 2001. During the fall semester 2004, she was a visiting Ph.D. student at Arizona State University. She has presented her work at several international conferences. The first study of this dissertation has recently been published in *Marketing Letters* and study 2 and 3 are forthcoming in the *Journal of Service Research* and *Marketing Letters*, respectively.

Currently, she works as an Assistant Professor at the Department of Marketing at Maastricht University. Her research interests are in the area of (online) marketing research, services marketing, and value creation by customers.